Submitted to the Annals of Applied Probability

A CENTRAL LIMIT THEOREM FOR ADAPTIVE AND INTERACTING MARKOV CHAINS

By G. Fort[*,†,‡], E. Moulines[†,§], P. Priouret[¶] and P. Vandekerkhove [‖]

CNRS & TELECOM ParisTech [‡§], *Univ. Pierre et Marie Curie*[¶], *Univ. Paris Est*[‖]

Adaptive and interacting Markov Chains Monte Carlo (MCMC) algorithms are a novel class of non-Markovian algorithms aimed at improving the simulation efficiency for complicated target distributions. In this paper, we study a general (non-Markovian) simulation framework covering both the adaptive and interacting MCMC algorithms. We establish a Central Limit Theorem for additive functionals of unbounded functions under a set of verifiable conditions, and identify the asymptotic variance. Our result extends all the results reported so far. An application to the interacting tempering algorithm (a simplified version of the equi-energy sampler) is presented to support our claims.

1. Introduction. Markov chain Monte Carlo (MCMC) methods generate samples from distributions known up to a scaling factor.

In the last decade, several non-Markovian simulation algorithms have been proposed. In the so-called adaptive MCMC algorithm, the transition kernel of the MCMC algorithm depends on a finite dimensional *parameter* which is updated at each iteration from the past values of the chain and the parameters. The prototypical example is the adaptive Metropolis algorithm, introduced in Haario et al. (1999) (see Saksman and Vihola (2010) and the references therein for recent references). Many other examples of adaptive MCMC algorithms are presented in the survey papers by Andrieu and Thoms (2008); Rosenthal (2009); Atchadé et al. (2011).

In the co-called *Interacting MCMC*, several processes are simulated in parallel, each targeting different distribution. Each process might interact with the whole past of its neighboring processes. A prototypical example is the equi-energy sampler introduced in Kou et al. (2006), where the different

[*]Corresponding author

[†]This work is partially supported by the French National Research Agency, under the program ANR-08-BLAN-0218 BigMC

AMS 2000 subject classifications: primary 65C05, 60F05, 62L10; secondary 65C40, 60J05,93E35

Keywords and phrases: MCMC, interacting MCMC, Limit theorems

processes target a tempered version of the target distribution. The convergence of this algorithm has been considered in a series of papers by Andrieu et al. (2007b), Andrieu et al. (2007a), Andrieu et al. (2011) and in Fort et al. (2010a). Different variants of the interacting MCMC algorithm have been later introduced and studied in Bercu et al. (2009), Del Moral and Doucet (2010) and Brockwell et al. (2010). These algorithms are so far limited to specific scenarios, and the assumptions used in these papers preclude the applications of their results in the applications considered in this paper.

The analysis of the convergence of these algorithms is involved. Whereas the basic building blocks of these simulation algorithms are Markov kernels, the processes generated by these techniques are no longer Markovian. Indeed, each individual process either interacts with its distant past, or the distant past of some auxiliary processes.

The ergodicity and the consistency of additive functionals for adaptive and interacting Markov Chains have been considered in several recent papers: see Fort et al. (2010a) and the references therein. Up to now, there are much fewer works addressing Central Limit Theorems (CLT). In Andrieu and Moulines (2006) the authors establish the asymptotic normality of additive functionals for a special class of adaptive MCMC algorithms in which a finite dimensional parameter is adapted using a stochastic approximation procedure. Some of the theoretical limitations of Andrieu and Moulines (2006) have been alleviated by Saksman and Vihola (2010) for the so-called adaptive Metropolis algorithm, which established a CLT for additive functionals for the Adaptive Metropolis algorithm (with a proof specially tailored for this algorithm). The results presented in this contribution contain as special cases these two earlier results.

The theory for interacting MCMC algorithms is up to now quite limited, despite the clear potential of this class of methods to sample complicated multimodal target distributions. The law of large numbers for additive functionals have been established in Andrieu et al. (2008) for some specific interacting algorithm. A wider class of interacting Markov chains has been considered in Del Moral and Doucet (2010). This paper establishes the consistency of a form of interacting tempering algorithm and provides non-asymptotic L^p-inequalities. The assumptions under which the results are derived are restrictive and the results do not cover the interacting MCMC algorithms considered in this paper. More recently, Fort et al. (2010a) have established the ergodicity and law of large numbers for a wide class of interacting MCMC, under the weakest conditions known so far.

A functional CLT was derived in Bercu et al. (2009) for a specific class of interacting Markov Chains but their assumptions do not cover the interactive

MCMC considered in this paper (and in particular, the interacting MCMC algorithm). A CLT for additive functionals is established by Atchadé (2010) for the interacting tempering algorithm; the proof of the main result in this paper, Theorem 3.3, contains a serious gap (p.865) which seems difficult to correct.

This paper aims at providing a theory removing the limitations mentioned above and covering both adaptive and interacting MCMC in a common unifying framework. The paper is organized as follows. In Section 2 we establish CLTs for adaptive and interacting MCMC algorithms. These results are applied in section 3 to the interacting tempering algorithm which is a simplified version of the Equi-Energy sampler. All the proofs are postponed in Section 4.

Notations. Let $(\mathsf{X}, \mathcal{X})$ be a general state space and P be a Markov transition kernel (see e.g. (Meyn and Tweedie, 2009, Chapter 3)). P acts on bounded functions f on X and on σ-finite positive measures μ on $\mathcal{X}$ via

$$Pf(x) \stackrel{\text{def}}{=} \int P(x, \mathrm{d}y) f(y) , \qquad \mu P(A) \stackrel{\text{def}}{=} \int \mu(\mathrm{d}x) P(x, A) .$$

We denote by P^n the n-iterated transition kernel defined inductively

$$P^n(x, A) \stackrel{\text{def}}{=} \int P^{n-1}(x, \mathrm{d}y) P(y, A) = \int P(x, \mathrm{d}y) P^{n-1}(y, A) ;$$

where P^0 is the identity kernel. For a function $V : \mathsf{X} \to [1, +\infty)$, define the V-norm of a function $f : \mathsf{X} \to \mathbb{R}$ by

$$|f|_V \stackrel{\text{def}}{=} \sup_{x \in \mathsf{X}} \frac{|f|(x)}{V(x)} .$$

When $V = 1$, the V-norm is the supremum norm denoted by $|f|_\infty$. Let $\mathcal{L}_V$ be the set of measurable functions such that $|f|_V < +\infty$. For μ a signed measure on $(\mathsf{X}, \mathcal{X})$, we defined $\|\mu\|_V$ the V-norm of μ as

$$\|\mu\|_V = \sup_{f \in \mathcal{L}_V, |f|_V \leq 1} |\mu(f)| .$$

When $V \equiv 1$, the V-norm corresponds to the total variation norm.

For two transition kernels P_1, P_2, define the V-distance as

$$\|P_1 - P_2\|_V \stackrel{\text{def}}{=} \sup_{x \in \mathsf{X}} V^{-1}(x) \left\| P_1(x, \cdot) - P_2(x, \cdot) \right\|_V .$$

Let $(x_n)_{n \in \mathbb{N}}$ a sequence. For $p \leq q \in \mathbb{N}^2$, $x_{p:q}$ denotes the vector $(x_p, \ldots, x_q)$.

2. Main results. Let $(\Theta, \mathcal{T})$ be a measurable space. Let $\{P_\theta, \theta \in \Theta\}$ be a collection of Markov transition kernels on $(\mathsf{X}, \mathcal{X})$ indexed by a parameter $\theta \in \Theta$. In the sequel, it is assumed that for any $A \in \mathcal{X}$, $(x, \theta) \mapsto P_\theta(x, A)$ is $\mathcal{X} \otimes \mathcal{T}/\mathcal{B}([0,1])$ measurable, where $\mathcal{B}([0,1])$ denotes the Borel σ-field. In the sequel Θ is not necessarily a finite-dimensional vector space. It might be a function space or a space of measures. We consider a $\mathsf{X} \times \Theta$-valued process $\{(X_n, \theta_n)\}_{n \in \mathbb{N}}$ on a filtered probability space $(\Omega, \mathcal{A}, \{\mathcal{F}_n, n \geq 0\}, \mathbb{P})$. It is assumed that

A1 The process $\{(X_n, \theta_n)\}_{n \in \mathbb{N}}$ is $(\mathcal{F}_n)_{n \in \mathbb{N}}$-adapted and for any bounded measurable function h,

$$\mathbb{E}\left[h(X_{n+1}) \,|\, \mathcal{F}_n\right] = P_{\theta_n} h(X_n) \; .$$

Assumption **A1** implies that conditional to the past (subsumed in the σ-algebra $\mathcal{F}_n$), the distribution of the next sample X_{n+1} is governed by the current value X_n and the current parameter θ_n. This assumption covers any adaptive and interacting MCMC algorithms; see Andrieu and Thoms (2008), Atchadé et al. (2011), Fort et al. (2010a) for examples. This assumption on the adaptation of the parameter $(\theta_n)_{n \in \mathbb{N}}$ is quite weak since it only requires the parameter to be adapted to the filtration. In practice, it frequently occurs that the joint process $\{(X_n, \theta_n)\}_{n \in \mathbb{N}}$ is Markovian but assumption **A1** covers more general adaptation rules.

We assume that the transition kernels $\{P_\theta, \theta \in \Theta\}$ satisfy a Lyapunov drift inequality and smallness conditions:

A2 For all $\theta \in \Theta$, P_θ is phi-irreducible, aperiodic and there exists a function $V : \mathsf{X} \to [1, +\infty)$, and for any $\theta \in \Theta$ there exist some constants $b_\theta \in (1, +\infty)$, $\lambda_\theta \in (0, 1)$ such that

$$P_\theta V \leq \lambda_\theta V + b_\theta \; .$$

In addition, for any $d \geq 1$ and any $\theta \in \Theta$, the level sets $\{V \leq d\}$ are 1-small for P_θ.

In many examples considered so far (see Andrieu and Moulines (2006), Saksman and Vihola (2010), Fort et al. (2010a), Andrieu et al. (2011)) this condition is satisfied. All the results below can be established under assumptions insuring that the drift inequality and/or the smallness condition are satisfied for some m-iterated P_θ^m. Note that checking assumption on the iterated kernel P_θ^m is prone to be difficult because the expression of the m-iterated kernel is most often rather involved.

A2 implies that, for any $\theta \in \Theta$, P_θ possesses an invariant probability distribution π_θ and the kernel P_θ is geometrically ergodic (Meyn and Tweedie, 2009, Chapter 15). The following lemma summarizes the properties of the family $\{P_\theta, \theta \in \Theta\}$ used in the sequel (see e.g. Douc et al. (2004) and references therein).

LEMMA 2.1. *Assume* ***A2***. *Then for any $\theta \in \Theta$, there exists a probability distribution π_θ such that $\pi_\theta P_\theta = \pi_\theta$ and $\pi_\theta(V) \leq b_\theta(1-\lambda_\theta)^{-1}$. In addition, for any $\alpha \in (0,1]$, the following property holds.*

> $\mathbf{P}[\alpha]$ *For any $\theta \in \Theta$, there exist $C_\theta < \infty$ and $\rho_\theta \in (0,1)$ such that, for any $\gamma \in [\alpha, 1]$,*
> $$\|P_\theta^n - \pi_\theta\|_{V^\gamma} \leq C_\theta \ \rho_\theta^n \ .$$

Set

$$L_\theta \stackrel{\text{def}}{=} C_\theta \vee (1-\rho_\theta)^{-1} \ . \tag{1}$$

It has been shown in Fort et al. (2010a), that under appropriate assumptions, when the sequence $(\theta_k)_{k\in\mathbb{N}}$ converges to $\theta_\star \in \Theta$ in an appropriate sense, $n^{-1}\sum_{k=1}^n f(X_k)$ converges almost surely to $\pi_{\theta_\star}(f)$, for any functions f belonging to a suitable class of functions $\mathcal{M}$.

The objective of this paper is to derive a CLT for $n^{-1/2}\sum_{k=1}^n \{f(X_k) - \pi_{\theta_\star}(f)\}$ for functions$f$ belonging to $\mathcal{M}$. To that goal, consider the following decomposition

$$n^{-1/2}\sum_{k=1}^n \{f(X_k) - \pi_{\theta_\star}(f)\} = S_n^{(1)}(f) + S_n^{(2)}(f) \ ,$$

where $S_n^{(1)}(f)$ and $S_n^{(2)}(f)$ are given by

$$S_n^{(1)}(f) \stackrel{\text{def}}{=} n^{-1/2}\sum_{k=1}^n \left\{f(X_k) - \pi_{\theta_{k-1}}(f)\right\} \ , \tag{2}$$

$$S_n^{(2)}(f) \stackrel{\text{def}}{=} n^{-1/2}\sum_{k=0}^{n-1} \{\pi_{\theta_k}(f) - \pi_{\theta_\star}(f)\} \ . \tag{3}$$

We consider these two terms separately. For the first term, we use a classical technique based on the Poisson decomposition; this amounts to write $S_n^{(1)}(f)$ as the sum of a martingale difference and of a remainder term converging to zero in probability; see Andrieu and Moulines (2006); Atchadé and Fort (2010); Fort et al. (2010a); Del Moral and Doucet (2010); Saksman and Vihola (2010) for law of large numbers for adaptive and interacting MCMC). Then we apply a classical CLT for martingale difference array; see for example (Hall and Heyde, 1980, Theorem 3.2).

The second term vanishes when $\pi_\theta = \pi_{\theta_\star}$ for all $\theta \in \Theta$ which is the case for example, for the adaptive Metropolis algorithm (Haario et al., 1999). In scenarios where $\theta \mapsto \pi_\theta$ is a non trivial function of θ, the weak convergence $S_n^{(2)}(f)$ relies on conditions which are quite problems specific.

The application detailed in Section 3, an elementary version of the interacting tempering algorithm, is a situation in which $\pi_{\theta_\star}$ is known but the expression of π_θ, $\theta \neq \theta_\star$, is unknown, except in very simple examples. The Wang-Landau algorithm (Wang and Landau, 2001; Liang et al., 2007) is an example of adaptive MCMC algorithm in which $\theta \mapsto \pi_\theta$ is explicit.

The results in this paper cover the case when the expression of π_θ is unknown: we rewrite $S_n^{(2)}(f)$ by using a linearization of the fluctuation $\pi_{\theta_k}(f) - \pi_{\theta_\star}(f)$ in terms of the difference $P_{\theta_k} - P_{\theta_\star}$

$$\pi_{\theta_k}(f) - \pi_{\theta_\star}(f) = \pi_{\theta_\star}\left(P_{\theta_k} - P_{\theta_\star}\right)\Lambda_{\theta_\star}(f) + \Xi(f, \theta_k) \; .$$

Our approach covers much more general set-up than the one outlined in Bercu et al. (2009).

By **A2**, for any $\alpha \in (0,1)$ and $f \in \mathcal{L}_{V^\alpha}$, the function $\sum_{n\geq 0} P_\theta^n\left(f - \pi_\theta(f)\right)$ exists and is in $\mathcal{L}_{V^\alpha}$. For $\theta \in \Theta$, denote by $\Lambda_\theta : \mathcal{L}_{V^\alpha} \mapsto \mathcal{L}_{V^\alpha}$ the transition kernel which associates to any function $f \in \mathcal{L}_{V^\alpha}$ the function $\Lambda_\theta f$ given by:

$$\Lambda_\theta f \stackrel{\text{def}}{=} \sum_{n\geq 0} P_\theta^n f - \pi_\theta(f) \; . \tag{4}$$

The function $\Lambda_\theta f$ is the solution of the Poisson equation

$$\Lambda_\theta f - P_\theta \Lambda_\theta f = f - \pi_\theta(f) \; . \tag{5}$$

This solution is unique up to an additive constant (see e.g. (Meyn and Tweedie, 2009, Proposition 17.4.1.)).

The convergence of $S_n^{(1)}(f)$ is addressed under the following assumptions which are related to the regularity in the parameter $\theta \in \Theta$ of the ergodic behavior of the kernels $\{P_\theta, \theta \in \Theta\}$.

A3 There exist $\alpha \in (0, 1/2)$ and a subset of measurable functions $\mathcal{M}_{V^\alpha} \subseteq \mathcal{L}_{V^\alpha}$ satisfying the two following conditions

(a) for any $f \in \mathcal{M}_{V^\alpha}$,

$$n^{-1/2} \sum_{k=1}^{n} \left|P_{\theta_k}\Lambda_{\theta_k} f - P_{\theta_{k-1}}\Lambda_{\theta_{k-1}} f\right|_{V^\alpha} V^\alpha(X_k) \xrightarrow{\mathbb{P}} 0 \; .$$

(b) $n^{-1/2\alpha} \sum_{k=0}^{n-1} L_{\theta_k}^{2/\alpha} P_{\theta_k} V(X_k) \xrightarrow{\mathbb{P}} 0$ where L_θ is defined by (1) for the constants C_θ, ρ_θ given by P[α].

A3-a controls the regularity in the parameter θ of the Poisson solution $\Lambda_\theta f$. Lemma 5.1 in Appendix 5 is useful to check **A3-a**. It relates the regularity in θ of the function $\theta \mapsto P_\theta \Lambda_\theta f$ to the ergodicity constants C_θ and ρ_θ introduced in Lemma 2.1 and to the regularity in θ of the function $\theta \mapsto P_\theta$ from the parameter space Θ to the space of Markov transition kernels equipped with the V-operator norm.

A3-b is a kind of containment condition (see Roberts and Rosenthal (2007)): when the ergodic behavior **A2** is uniform in θ so that λ_θ, b_θ and the minorization constant of the P_θ-smallness condition do not depend on θ, then the constant L_θ does not depend on θ and by **A1** and the drift inequality **A2**,

$$n^{-1/2\alpha} \sum_{k=0}^{n-1} \mathbb{E}\left[V(X_{k+1})\right] \leq n^{1-1/2\alpha} \ \left\{\mathbb{E}\left[V(X_0)\right] + (1-\lambda)^{-1} b\right\} \to 0 \ .$$

Therefore, condition **A3-b** holds provided the ergodic constant L_{θ_k} is controlled by a slowly-increasing function of k. Lemma 5.2 in Appendix 5 provides sufficient conditions to obtain upper bounds of $\theta \mapsto L_\theta$ in terms of the constants appearing in the drift inequality **A2**.

We finally introduce a condition allowing to obtain a closed-form expression for the asymptotic variance of $S_n^{(1)}(f)$. For $\theta \in \Theta$ and $f \in \mathcal{L}_{V^\alpha}$ define

$$F_\theta \stackrel{\text{def}}{=} P_\theta (\Lambda_\theta f)^2 - \left[P_\theta \Lambda_\theta f\right]^2 \ . \tag{6}$$

A4 For any $f \in \mathcal{M}_{V^\alpha}$, $n^{-1} \sum_{k=0}^{n-1} F_{\theta_k}(X_k) \xrightarrow{\mathbb{P}} \sigma^2(f)$, where $\sigma^2(f)$ is a deterministic constant.

Assumption **A4** is typically established by using the Law of Large Numbers (LLN) for adaptive and interacting Markov Chain derived in Fort et al. (2010a); see also Theorem 5.4 in Appendix 5. Under appropriate regularity conditions on the Markov kernels $\{P_\theta, \theta \in \Theta\}$, it is proved that $n^{-1} \sum_{k=0}^{n-1} \{F_{\theta_k}(X_k) - \int \pi_{\theta_k}(\mathrm{d}x)\, F_{\theta_k}(x)\}$ converges in probability to zero. The second step consists in showing that $n^{-1} \sum_{k=0}^{n-1} \int \pi_{\theta_k}(\mathrm{d}x)\, F_{\theta_k}(x)$ converges to a (deterministic) constant $\sigma^2(f)$: when π_θ is not explicitly known and the set X is Polish, Lemma 5.3 in Appendix 5 is useful to check this convergence. In practice, this may introduce a restriction of the set of functions $f \in \mathcal{L}_{V^\alpha}$ for which this limit holds (see e.g. the example detailed in Section 3 where $\mathcal{M}_{V^\alpha} \neq \mathcal{L}_{V^\alpha}$).

We can now state conditions upon which $S_n^{(1)}(f)$ is asymptotically normal.

THEOREM 2.2. *Assume **A1** to **A4**. For any $f \in \mathcal{M}_{V^\alpha}$,*

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \left\{ f(X_k) - \pi_{\theta_{k-1}}(f) \right\} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(f)) \; .$$

The proof is in section 4.1.1. When $\pi_\theta = \pi$ for any θ, Theorem 2.2 provides sufficient conditions for a CLT for additive functionals to hold.

When π_θ is a function of $\theta \in \Theta$, we need now to obtain a joint CLT for $(S_n^{(1)}(f), S_n^{(2)}(f))$ (see (2) and (3)). To that goal, we replace **A1** by the following assumption which implies that, conditionally to the process $(\theta_k)_{k\in\mathbb{N}}$, $(X_k)_{k\in\mathbb{N}}$ is an inhomogeneous Markov chain with transition kernels $(P_{\theta_j}, j \geq 0)$:

A5 There exists an initial distribution ν such that for any bounded measurable function $f : \mathsf{X}^{n+1} \to \mathbb{R}$,

$$\mathbb{E}\left[f(X_{0:n}) \,|\, \theta_{0:n} \right] = \int \cdots \int \nu(\mathrm{d}x_0) f\left(x_{0:n}\right) \prod_{j=1}^{n} P_{\theta_{j-1}}(x_{j-1}, \mathrm{d}x_j) \; .$$

Assumption **A5** is satisfied when $\{(X_n, \theta_n)\}_{n\in\mathbb{N}}$ is an interacting MCMC algorithm. Note that **A5** implies **A1**.

The first step in the proof of the joint CLT consists in linearizing the difference $\pi_{\theta_n} - \pi_{\theta_\star}$. Under **A2**, $\pi_\theta(g)$ exists for any $g \in \mathcal{L}_{V^\alpha}$ and $\theta \in \Theta$ (see Lemma 2.1), and we have

$$\pi_\theta(g) - \pi_{\theta_\star}(g) = \pi_\theta P_\theta g - \pi_{\theta_\star} P_{\theta_\star} g = \pi_\theta \left(P_\theta - P_{\theta_\star}\right) g + \left(\pi_\theta - \pi_{\theta_\star}\right) P_{\theta_\star} g \; ,$$

which implies that $\left(\pi_\theta - \pi_{\theta_\star}\right)\left(\mathrm{I} - P_{\theta_\star}\right) g = \pi_\theta \left(P_\theta - P_{\theta_\star}\right) g$. Let $f \in \mathcal{L}_{V^\alpha}$. Then $\Lambda_{\theta_\star} f \in \mathcal{L}_{V^\alpha}$ and by applying the previous equality with $g = \Lambda_{\theta_\star} f$, we have by (5)

$$\pi_\theta(f) - \pi_{\theta_\star}(f) = \pi_\theta \left(P_\theta - P_{\theta_\star}\right) \Lambda_{\theta_\star} f \; . \tag{7}$$

We can iterate this decomposition, writing

$$\begin{aligned} \pi_\theta(f) - \pi_{\theta_\star}(f) = \pi_{\theta_\star} &\left(P_\theta - P_{\theta_\star}\right) \Lambda_{\theta_\star} f + \\ &\pi_\theta \left(\left(P_\theta - P_{\theta_\star}\right) \Lambda_{\theta_\star} f\right) - \pi_{\theta_\star} \left(\left(P_\theta - P_{\theta_\star}\right) \Lambda_{\theta_\star} f\right) \end{aligned}$$

Applying again (7), we obtain

$$\pi_\theta(f) - \pi_{\theta_\star}(f) = \pi_{\theta_\star} \left(P_\theta - P_{\theta_\star}\right) \Lambda_{\theta_\star} f + \pi_\theta \left(P_\theta - P_{\theta_\star}\right) \Lambda_{\theta_\star} \left(P_\theta - P_{\theta_\star}\right) \Lambda_{\theta_\star} f \; .$$

This decomposition can be iterated, which yields The first term in the RHS of the previous equation is the leading term of the error $\pi_{\theta_k} - \pi_{\theta_\star}$, whereas the second term is a remainder. This decomposition naturally leads to the following assumption.

A6 For any function $f \in \mathcal{M}_{V^\alpha}$,

(a) there exists a positive constant $\gamma^2(f)$ such that

$$n^{-1/2} \sum_{k=1}^{n} \pi_{\theta_\star} \left(P_{\theta_k} - P_{\theta_\star}\right) \Lambda_{\theta_\star} f \xrightarrow{\mathcal{D}} \mathcal{N}(0, \gamma^2(f)) \,. \tag{8}$$

(b) $n^{-1/2} \sum_{k=1}^{n} \pi_{\theta_k} \left(P_{\theta_k} - P_{\theta_\star}\right) \Lambda_{\theta_\star} \left(P_{\theta_k} - P_{\theta_\star}\right) \Lambda_{\theta_\star} f \xrightarrow{\mathbb{P}} 0.$

THEOREM 2.3. *Assume **A2** to A6. For any function $f \in \mathcal{M}_{V^\alpha}$,*

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \{f(X_k) - \pi_{\theta_\star}(f)\} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma^2(f) + \gamma^2(f)\right) \,.$$

The proof of Theorem 2.3 is postponed to section 4.1.2. It is worthwhile to note that, as a consequence of **A5**, the variance is additive. This result extends Bercu et al. (2009) which addresses the case when $P_\theta(x, A) = P_\theta(A)$ *i.e.* the case when conditionally to the adaptation process $(\theta_n)_{n\in\mathbb{N}}$, the random variables $(X_n)_{n\in\mathbb{N}}$ are independent (see (Bercu et al., 2009, Eq. (1.4))). Our result, applied in this simpler situation, yields to the same asymptotic variance.

3. Application to Interacting Tempering algorithm.

We consider the simplified version of the equi-energy sampler (Kou et al., 2006) introduced in Andrieu et al. (2011). This version is referred to as the Interacting-tempering (IT) sampler. Recently, convergence of the marginals and strong law of large numbers results have been established under general conditions (see Fort et al. (2010a)). In this section, we derive a CLT under similar assumptions.

Let $\{\pi^{\beta_k}, k \in \{1, \cdots, K\}\}$ be a sequence of tempered densities on X, where $0 < \beta_1 < \cdots < \beta_K = 1$. At the first level, a process $(Y_k)_{k\in\mathbb{N}}$ with stationary distribution proportional to π^{β_1} is run. At the second level, a process $(X_k)_{k\in\mathbb{N}}$ with stationary distribution proportional to π^{β_2} is constructed: at each iteration the next value is obtained from a Markov kernel depending on the occupation measure of the chain $(Y_k)_{k\in\mathbb{N}}$ up to the current time-step. This 2-stages mechanism is then repeated to design a process targeting π^{β_k} by using the occupation measure of the process targeting $\pi^{\beta_{k-1}}$.

For ease of exposition, it is assumed that $(\mathsf{X}, \mathcal{X})$ is a Polish space equipped with its Borel σ-field, and the densities are w.r.t. some σ-finite measure on $(\mathsf{X}, \mathcal{X})$. We address the case $K = 2$ and discuss below possible extensions to the case $K > 2$.

We start with a description of the IT (case $K = 2$). Denote by Θ the set of the probability measures on $(\mathsf{X}, \mathcal{X})$ equipped with the Borel sigma-field $\mathcal{T}$ associated to the topology of weak convergence. Let P be a transition kernel on $(\mathsf{X}, \mathcal{X})$ with unique invariant distribution π (typically, P is chosen to be a Metropolis-Hastings kernel). Denote by $\epsilon \in (0, 1)$ the probability of interaction. Let $(Y_k)_{k\in\mathbb{N}}$ be a discrete-time (possibly non-stationary) process and denote by θ_n the empirical probability measure:

$$\theta_n \stackrel{\text{def}}{=} \frac{1}{n} \sum_{k=1}^{n} \delta_{Y_k} \; . \tag{9}$$

Choose $X_0 \sim \nu$. At the n-th iteration of the algorithm, two actions may be taken:

1. with probability $(1 - \epsilon)$, the state X_{n+1} is sampled from the Markov kernel $P(X_n, \cdot)$,
2. with probability ϵ, a tentative state Z_{n+1} is drawn uniformly from the past of the auxiliary process $\{Y_k, k \leq n\}$. This move is accepted with probability $r(X_n, Z_{n+1})$, where the acceptance ratio r is given by

$$r(x, z) \stackrel{\text{def}}{=} 1 \wedge \frac{\pi(z)\pi^{1-\beta}(x)}{\pi^{1-\beta}(z)\pi(x)} = 1 \wedge \frac{\pi^{\beta}(z)}{\pi^{\beta}(x)} \; . \tag{10}$$

Define the family of Markov transition kernels $\{P_\theta, \theta \in \Theta\}$ by

$$\begin{aligned} P_\theta(x, A) \stackrel{\text{def}}{=}& (1 - \epsilon)P(x, A) \\ &+ \epsilon \left(\int_A r(x, y)\theta(\mathrm{d}y) + \mathbb{1}_A(x) \int \{1 - r(x, y)\}\, \theta(\mathrm{d}y) \right) \; . \end{aligned} \tag{11}$$

Then, the above algorithmic description implies that the bivariate process $\{(X_n, \theta_n)\}_{n\in\mathbb{N}}$ is such that for any bounded function h on X^{n+1}

$$\mathbb{E}\left[h(X_{0:n})|\theta_{0:n}\right] = \int \nu(\mathrm{d}x_0) P_{\theta_0}(x_0, \mathrm{d}x_1) \cdots P_{\theta_{n-1}}(x_{n-1}, \mathrm{d}x_n)\, h(x_{0:n}) \; .$$

We apply the results of Section 2 in order to prove that the IT process $(X_k)_{k\in\mathbb{N}}$ satisfies a CLT. To that goal, it is assumed that the target density π and the transition kernel P satisfy the following conditions:

I1 π is a continuous positive density on X and $|\pi|_\infty < +\infty$.

I2 (a) P is a phi-irreducible aperiodic Feller transition kernel on $(\mathsf{X}, \mathcal{X})$ such that $\pi P = \pi$.

(b) There exist $\tau \in (0,1)$, $\lambda \in (0,1)$ and $b < +\infty$ such that

$$PV \leq \lambda V + b \quad \text{with} \quad V(x) \stackrel{\text{def}}{=} (\pi(x)/|\pi|_\infty)^{-\tau} \ . \tag{12}$$

(c) For any $p \in (0, |\pi|_\infty)$, the sets $\{\pi \geq p\}$ are 1-small (w.r.t. the transition kernel P).

(d) For any $\gamma \in (0, 1/2)$ and any equicontinuous set of functions $\mathcal{F} \subseteq \mathcal{L}_{V^\gamma}$, the set of functions $\{Ph : h \in \mathcal{F}, |h|_{V^\gamma} \leq 1\}$ is equicontinuous.

From the expression of the acceptance ratio r (see Eq. (10)) and the assumption I2-a, it holds

$$\pi P_{\theta_\star} = \pi \ ,$$

where $\theta_\star \propto \pi^{1-\beta}$. Therefore, when θ_n converges to $\theta_\star$, it is expected that $(X_k)_{k \in \mathbb{N}}$ behaves asymptotically as π; see Fort et al. (2010a).

Drift conditions for the symmetric random walk Metropolis (SRWM) algorithm are discussed in Roberts and Tweedie (1996), Jarner and Hansen (2000) and Saksman and Vihola (2010). Under conditions which imply that the target density π is super-exponential in the tails and have regular contours, Jarner and Hansen (2000) and Saksman and Vihola (2010) show that any functions proportional to π^{-s} with $s \in (0,1)$ satisfies a Foster-Lyapunov drift inequality (Jarner and Hansen, 2000, Theorems 4.1 and 4.3). Under this condition, I2-b is satisfied with any τ in the interval $(0,1)$. Assumption I2-d holds for the SRWM kernel under weak conditions on the symmetric proposal distribution as shown by the following lemma. The proof is in section 4.2.1.

LEMMA 3.1. *Assume I1. Let P be a Metropolis kernel with invariant distribution π and a symmetric proposal distribution $q : \mathsf{X} \times \mathsf{X} \to \mathbb{R}^+$ such that $\sup_{(x,y) \in \mathsf{X}^2} q(x,y) < +\infty$ and the function $x \mapsto q(x, \cdot)$ is continuous from $(\mathsf{X}, |\cdot|)$ to the set of probability densities equipped with the total variation norm. Then P satisfies I2-d with any function $V \propto \pi^{-\tau}$, $\tau \in [0,1)$, such that $\pi(V) < +\infty$.*

For a measurable function $f : \mathsf{X} \to \mathbb{R}$ such that $\theta_\star(|f|) < +\infty$, define the following sequence of random processes on $[0,1]$:

$$t \mapsto S_n(f;t) = n^{-1/2} \sum_{j=1}^{\lfloor nt \rfloor} \{f(Y_j) - \theta_\star(f)\} \ . \tag{13}$$

It is assumed that the auxiliary process $\{Y_n \ , n \geq 0\}$ converges to the probability distribution $\theta_\star$ in the following sense:

I3 (a) $\theta_\star(V) < +\infty$ and $\sup_n \mathbb{E}\left[V(Y_n)\right] < +\infty$.

(b) There exists a space $\mathcal{N}$ of real-valued measurable functions defined on X such that $V \in \mathcal{N}$ and for any function $f \in \mathcal{N}$, $\theta_n(f) \xrightarrow{\text{a.s.}} \theta_\star(f)$.

(c) For any function $f \in \mathcal{N}$, the sequence of processes $(S_n(f,t), n \geq 1, t \in [0,1])$ converges in distribution to $(\tilde{\gamma}(f)B(t), t \in [0,1])$, where $\tilde{\gamma}(f)$ is a non-negative constant and $(B(t) : t \in [0,1])$ is a standard Brownian motion.

(d) For any $\alpha \in (0, 1/2)$, there exist constants ϱ_0 and ϱ_1 such that, for any integers $n, k \geq 1$, for any measurable function $h : \mathsf{X}^k \to \mathbb{R}$ satisfying $|h(y_1, \ldots, y_k)| \leq \sum_{j=1}^k V^\alpha(y_j)$,

$$\mathbb{E}\left(\int \cdots \int \prod_{j=1}^k \left[\theta_n(\mathrm{d}y_j) - \theta_\star(\mathrm{d}y_j)\right] h(y_1, \ldots, y_k)\right)^2 \leq A_k\, n^{-k} ,$$

with $\limsup_k \ln A_k / (k \ln k) < \infty$.

I3 is satisfied when $(Y_k)_{k\in\mathbb{N}}$ is i.i.d. with distribution $\theta_\star$ such that $\theta_\star(V) < +\infty$. In that case, I3-b to I3-c hold for any measurable function f such that $\theta_\star(|f|^2) < +\infty$. I3-d is satisfied using (Serfling, 1980, Lemma A, pp. 190).

I3 is also satisfied when $(Y_k)_{k\in\mathbb{N}}$ is an asymptotically stationary Markov chain with transition kernel Q. In that case, I3-a to I3-c are satisfied for any measurable function f such that $\theta_\star\left(|f\left[(I - Q)^{-1} f\right]|\right) < +\infty$ (see e.g. (Meyn and Tweedie, 2009, Chapter 17)). Condition I3-d for a (non-stationary) geometrically ergodic Markov chain is established in the supplementary paper (Fort et al., 2011).

The following proposition shows that under I1 and I2, condition **A2** holds with the drift function V given by A2-b. It also provides a control of the ergodicity constants C_θ, ρ_θ in Lemma 2.1. The proof is a direct consequence of (Fort et al., 2010a, Proposition 3.1, Corollary 3.2), Lemmas 2.1 and 5.2, and is omitted.

PROPOSITION 3.2. *Assume I1 and I2a-b-c. For any $\theta \in \Theta$, P_θ is phi-irreducible, aperiodic. In addition, there exist $\tilde{\lambda} \in (0,1)$ and $\tilde{b} < +\infty$ such that, for any $\theta \in \Theta$,*

$$P_\theta V(x) \leq \tilde{\lambda} V(x) + \tilde{b}\, \theta(V) , \quad \text{for all } x \in \mathsf{X}. \tag{14}$$

The property $P[\alpha]$ holds for any $\alpha \in (0, 1/2)$, and there exists C such that for any $\theta \in \Theta$, $L_\theta \leq C\theta(V)$.

Assume in addition I3a and $\mathbb{E}[V(X_0)] < +\infty$. Then, $\sup_{n\geq 0} \mathbb{E}\left[V(X_n)\right] < +\infty$.

The next step is to check assumptions A3 and **A4**.

PROPOSITION 3.3. *Assume I1, I2, I3a-b and* $\mathbb{E}[V(X_0)] < +\infty$*. For any* $\alpha \in (0, 1/2)$*, set* $\mathcal{M}_{V^\alpha}$ *be the set of continuous functions belonging to* $\mathcal{L}_{V^\alpha} \cap \mathcal{N}$*. Then, for any* $\alpha \in (0, 1/2)$*, the conditions A3 and* ***A4*** *hold with*

$$\sigma^2(f) \stackrel{\text{def}}{=} \int \pi_{\theta_\star}(\mathrm{d}x) F_{\theta_\star}(x) \ , \tag{15}$$

where F_θ *is given by (6).*

The proof is postponed to Appendix 4.2.2. We can now apply Theorem 2.3 and prove a CLT for the 2-levels IT.

THEOREM 3.4. *Assume I1, I2, I3 and* $\mathbb{E}[V(X_0)] < +\infty$*. Then, for any* $\alpha \in (0, 1/2)$ *and any continuous function* $f \in \mathcal{L}_{V^\alpha} \cap \mathcal{N}$*,*

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} (f(X_k) - \pi_{\theta_\star}(f)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(f) + 2\tilde{\gamma}^2(f)) \ ,$$

where $\sigma^2(f)$ *and* $\tilde{\gamma}^2(f)$ *are given by (15) and I3-c.*

The proof is postponed to Appendix 4.2.3.

The above discussion could be repeated in order to prove by induction a CLT for the K-level IT when $K > 2$ (see Fort et al. (2010a) for a similar approach in the proof of the ergodicity and the LLN for IT). Nevertheless, the main difficulty is to iterate the control of the L^2-moment for the V-statistics (see I3-d) when $(Y_k)_{k\in\mathbb{N}}$ is not a Markov Chain or, more generally, a process satisfying some mixing conditions. A similar difficulty has been reported in Andrieu et al. (2011).

4. Proofs.

Denote by $D_V(\theta, \theta')$ the V-distance of the kernels P_θ and $P_{\theta'}$:

$$D_V(\theta, \theta') \stackrel{\text{def}}{=} \|P_\theta - P_{\theta'}\|_V \ . \tag{16}$$

Note that under **A2**, for any $\alpha \in (0, 1]$, any $f \in \mathcal{L}_{V^\alpha}$ and any $\theta \in \Theta$,

$$|\Lambda_\theta f|_{V^\alpha} \leq |f|_{V^\alpha} \ L_\theta^2 \tag{17}$$

where L_θ is defined by (1).

4.1. *Proofs of the results in Section 2.*

4.1.1. *Proof of Theorem 2.2.* Let $f \in \mathcal{M}_{V^\alpha}$. Eq. (5) yields to $S_n^{(1)}(f) = \Xi_n(f) + R_n^{(1)}(f) + R_n^{(2)}(f)$ with

$$\begin{aligned}
\Xi_n(f) &\stackrel{\text{def}}{=} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \{\Lambda_{\theta_{k-1}} f(X_k) - P_{\theta_{k-1}} \Lambda_{\theta_{k-1}} f(X_{k-1})\} , \\
R_n^{(1)}(f) &\stackrel{\text{def}}{=} n^{-1/2} \sum_{k=1}^{n} \{P_{\theta_k} \Lambda_{\theta_k} f(X_k) - P_{\theta_{k-1}} \Lambda_{\theta_{k-1}} f(X_k)\} , \\
R_n^{(2)}(f) &\stackrel{\text{def}}{=} n^{-1/2} P_{\theta_0} \Lambda_{\theta_0} f(X_0) - n^{-1/2} P_{\theta_n} \Lambda_{\theta_n} f(X_n) .
\end{aligned}$$

We first show that the two remainders terms $R_n^{(1)}(f)$ and $R_n^{(2)}(f)$ converge to zero in probability. We have

$$|P_\theta \Lambda_\theta f(x) - P_{\theta'} \Lambda_{\theta'} f(x)| \leq |P_\theta \Lambda_\theta f(x) - P_{\theta'} \Lambda_{\theta'} f(x)|_{V^\alpha} \, V^\alpha(x) .$$

Assumption A3 implies that $R_n^{(1)}(f)$ converges to zero in probability. The drift inequality **A2** combined with the Jensen's inequality imply $P_\theta V^\alpha \leq \lambda_\theta^\alpha V^\alpha + b_\theta^\alpha$. By (17) and this inequality,

$$|P_\theta \Lambda_\theta f(x)| \leq |f|_{V^\alpha} \; L_\theta^2 \, P_\theta V^\alpha(x) \leq |f|_{V^\alpha} \; L_\theta^2 \, (V^\alpha(x) + b_\theta^\alpha) .$$

Then, $P_{\theta_0} \Lambda_{\theta_0} f(X_0)$ is finite w.p.1. and $n^{-1/2} P_{\theta_0} \Lambda_{\theta_0} f(X_0) \xrightarrow{\text{a.s.}} 0$. By **A3-b** and (17), $n^{-1/2} P_{\theta_n} \Lambda_{\theta_n} f(X_n) \xrightarrow{\mathbb{P}} 0$. Hence, $R_n^{(2)}(f) \xrightarrow{\mathbb{P}} 0$.

We now consider $\Xi_n(f)$. Set $D_k(f) \stackrel{\text{def}}{=} \Lambda_{\theta_{k-1}} f(X_k) - P_{\theta_{k-1}} \Lambda_{\theta_{k-1}} f(X_{k-1})$. Observe that under **A1**, $D_k(f)$ is a martingale-increment w.r.t. the filtration $\{\mathcal{F}_k, k \geq 0\}$. The limiting distribution for $\Xi_n(f)$ follows from martingale CLT (see *e.g.* (Hall and Heyde, 1980, Corollary 3.1.)). We check the conditional Lindeberg condition. Let $\epsilon > 0$. Under **A2**, we have by (17)

$$D_k(f) \leq |f|_{V^\alpha} \; \left| L_{\theta_{k-1}}^2 \; \left\{ V^\alpha(X_k) + P_{\theta_{k-1}} V^\alpha(X_{k-1}) \right\} \right| .$$

Set $\tau \stackrel{\text{def}}{=} 1/\alpha - 2 > 0$.

$$\begin{aligned}
&\frac{1}{n} \sum_{k=1}^{n} \mathbb{E}\left[D_k^2(f) \mathbb{1}_{|D_k(f)| \geq \epsilon\sqrt{n}} \,\middle|\, \mathcal{F}_{k-1} \right] \leq \left(\frac{1}{\epsilon\sqrt{n}} \right)^\tau \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}\left[D_k^{2+\tau}(f) \,\middle|\, \mathcal{F}_{k-1} \right] \\
&\leq |f|_{V^\alpha}^{2+\tau} \; \left(\frac{1}{\epsilon\sqrt{n}} \right)^\tau \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}\left[L_{\theta_{k-1}}^{2(2+\tau)} \; \left\{ V^\alpha(X_k) + P_{\theta_{k-1}} V^\alpha(X_{k-1}) \right\}^{2+\tau} \,\middle|\, \mathcal{F}_{k-1} \right] \\
&\leq 2^{2+\tau} \, |f|_{V^\alpha}^{2+\tau} \; \left(\frac{1}{\epsilon\sqrt{n}} \right)^\tau \frac{1}{n} \sum_{k=0}^{n-1} L_{\theta_k}^{2(2+\tau)} \, P_{\theta_k} V(X_k) .
\end{aligned}$$

Under **A3-b**, the RHS converges to zero in probability thus concluding the proof of the conditional Lindeberg condition. For the limiting variance condition, observe that

$$\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\left[D_k^2(f)\,\middle|\,\mathcal{F}_{k-1}\right] = \frac{1}{n}\sum_{k=0}^{n-1}F_{\theta_k}(X_k)\;,$$

where F_θ is given by (6) and, under **A4**, $n^{-1}\sum_{k=1}^{n}\mathbb{E}\left[D_k^2\,\middle|\,\mathcal{F}_{k-1}\right] \xrightarrow{\mathbb{P}} \sigma^2(f)$. This concludes the proof.

4.1.2. *Proof of Theorem 2.3.* We start by establishing a joint CLT for $(S_n^{(1)}(f), S_n^{(2)}(f))$, where $S_n^{(1)}(f)$ and $S_n^{(2)}(f)$ are defined in (2) and (3), respectively. Similar to the proof of Theorem 2.2, we write $S_n^{(1)}(f) = \Xi_n(f) + R_n^{(1)}(f) + R_n^{(2)}(f)$ and prove that $R_n^{(1)}(f) + R_n^{(2)}(f) \xrightarrow{\mathbb{P}} 0$. We thus consider the convergence of $\Xi_n(f) + S_n^{(2)}(f)$. Set $\mathcal{F}_n^\theta \stackrel{\text{def}}{=} \sigma(\theta_k, k \le n)$. Under **A5**,

$$\mathbb{E}\left[\mathrm{e}^{\mathrm{i}(u_1\Xi_n(f)+u_2S_n^{(2)}(f))}\right] = \mathbb{E}\left[\mathbb{E}\left[\mathrm{e}^{\mathrm{i}u_1\Xi_n(f)}\,\middle|\,\mathcal{F}_n^\theta\right]\mathrm{e}^{\mathrm{i}u_2S_n^{(2)}(f)}\right]\;.$$

Applying the conditional CLT (Douc and Moulines, 2008, Theorem A.3.) with the filtration $\mathcal{F}_{n,k} \stackrel{\text{def}}{=} \sigma(Y_1,\cdots,Y_n,X_1,\cdots,X_k)$, yields to:

$$\lim_{n\to\infty}\mathbb{E}\left[\mathrm{e}^{\mathrm{i}u_1\Xi_n(f)}\,\middle|\,\mathcal{F}_n^\theta\right] \xrightarrow{\mathbb{P}} \mathrm{e}^{-u_1^2\sigma^2(f)/2}\;; \tag{18}$$

observe that under **A5**, the conditions (31) and (32) in Douc and Moulines (2008) can be proved following the same lines as in the proof of Theorem 2.2; details are omitted. Therefore,

$$\begin{aligned}\mathbb{E}\left[\mathrm{e}^{\mathrm{i}(u_1\Xi_n(f)+u_2S_n^{(2)}(f))}\right] = \mathbb{E}&\left[\left(\mathbb{E}\left[\mathrm{e}^{\mathrm{i}u_1\Xi_n(f)}\,\middle|\,\mathcal{F}_n^\theta\right] - \mathrm{e}^{-u_1^2\sigma^2(f)/2}\right)\mathrm{e}^{iu_2S_n^{(2)}(f)}\right]\\ &+ \mathrm{e}^{-u_1\sigma^2(f)/2}\mathbb{E}\left[\mathrm{e}^{\mathrm{i}u_2S_n^{(2)}(f)}\right]\;.\end{aligned}$$

By (18), the first term in the RHS of the previous equation converges to zero. Under **A6**, $\lim_{n\to\infty}\mathbb{E}\left[\mathrm{e}^{\mathrm{i}u_2S_n^{(2)}(f)}\right] = \mathrm{e}^{-u_2^2\gamma^2(f)/2}$ and this concludes the proof.

4.2. *Proofs of Section 3.*

4.2.1. *Proof of Lemma 3.1.* Let $\gamma \in (0,1/2)$ and $\mathcal{F}$ be an equicontinuous set of functions in $\mathcal{L}_{V^\gamma}$. Let $h \in \mathcal{F}$, $|h|_{V^\gamma} \leq 1$. By construction, the transition kernel of a symmetric random walk Metropolis with proposal transition density $q(x,\cdot)$ and target density π may be expressed as

$$Ph(x) = \int r(x,y)h(y)q(x,y)\,\mathrm{d}y + h(x)\int \{1 - r(x,y)\}\, q(x,y)\mathrm{d}y \;,$$

where $r(x,y) \stackrel{\mathrm{def}}{=} 1 \wedge (\pi(y)/\pi(x))$ is the acceptance ratio. Therefore, the difference $Ph(x) - Ph(x')$ may be bounded by

$$\begin{aligned}\left|Ph(x) - Ph(x')\right| &\leq 2\left|h(x) - h(x')\right| \\ &+ \int \left|h(y) - h(x')\right| \; \left|r(x,y) - r(x',y)\right| q(x,y)\mathrm{d}y \\ &+ \left|\int \left(h(y) - h(x')\right) \; r(x',y)\left(q(x,y) - q(x',y)\right)\mathrm{d}y\right| \;.\end{aligned}$$

Since $|r(x,y) - r(x',y)| \leq \pi(y)|\pi^{-1}(x) - \pi^{-1}(x')|$,

$$\begin{aligned}&\int \left|h(y) - h(x')\right| \; \left|r(x,y) - r(x',y)\right| q(x,y)\mathrm{d}y \\ &\quad \leq \left|\pi^{-1}(x) - \pi^{-1}(x')\right| \int \left|h(y) - h(x')\right| \pi(y) \; q(x,y)\;\mathrm{d}y \\ &\quad \leq \left(\sup_{(x,y)\in\mathsf{X}^2} q(x,y)\right) \; \left|\pi^{-1}(x) - \pi^{-1}(x')\right| \; \left(\pi(V^\gamma) + V^\gamma(x')\right) \;.\end{aligned}$$

In addition,

$$\begin{aligned}&\left|\int \left(h(y) - h(x')\right) \; r(x',y)\left(q(x,y) - q(x',y)\right)\mathrm{d}y\right| \\ &= \left|\int_{\{y:\pi(y)\leq\pi(x')\}} \left(h(y) - h(x')\right) \; \frac{\pi(y)}{\pi(x')}\left(q(x,y) - q(x',y)\right)\mathrm{d}y\right| \\ &+ \left|\int_{\{y:\pi(y)>\pi(x')\}} \left(h(y) - h(x')\right) \; \left(q(x,y) - q(x',y)\right)\mathrm{d}y\right| \\ &\leq 4\;\pi^{-1}(x') \; \left\|q(x,\cdot) - q(x',\cdot)\right\|_{\mathrm{TV}} \; \sup_{y\in\mathsf{X}}|h(y)\;\pi(y)| \;.\end{aligned}$$

Since $V \propto \pi^{-\tau}$ and $\tau \in (0,1)$, $\sup_{\mathsf{X}} |h|\pi \leq 1$ under I1. Therefore, there exists a constant C such that for any $h \in \{h \in \mathcal{F}, |h|_{V^\gamma} \leq 1\}$ and any $x, x' \in \mathsf{X}$,

$$\begin{aligned}&\left|Ph(x) - Ph(x')\right| \leq 2\left|h(x) - h(x')\right| \\ &\quad + C\left(\left|\pi^{-1}(x) - \pi^{-1}(x')\right| + \left\|q(x,\cdot) - q(x',\cdot)\right\|_{\mathrm{TV}}\right)\left(V^\gamma(x') + \pi^{-1}(x')\right) \;,\end{aligned}$$

thus concluding the proof.

4.2.2. *Proof of Proposition 3.3.* The proof is prefaced by several lemmas.

LEMMA 4.1. *Let P_θ be the transition kernel given by* (11). *For any $(\theta, \theta') \in \Theta^2$ and any $\alpha \in (0,1]$, we have $D_{V^\alpha}(\theta, \theta') \leq 2\|\theta - \theta'\|_{V^\alpha}$. For any positive integer n,*

$$D_{V^\alpha}(\theta_n, \theta_{n-1}) \leq \frac{2}{n}\theta_{n-1}(V^\alpha) + \frac{2}{n}V^\alpha(Y_n) \;. \tag{19}$$

PROOF. For any $f \in \mathcal{L}_{V^\alpha}$ and any $x \in \mathsf{X}$,

$$\begin{aligned} |P_\theta(x,f) - P_{\theta'}(x,f)| &= \epsilon \left| \int r(x,y)\,[f(y) - f(x)]\,\left[\theta(\mathrm{d}y) - \theta'(\mathrm{d}y)\right] \right| \\ &\leq \epsilon \left\|\theta - \theta'\right\|_{V^\alpha} |r(x,\cdot)\,[f(\cdot) - f(x)]|_{V^\alpha} \leq 2\left\|\theta - \theta'\right\|_{V^\alpha} |f|_{V^\alpha} \;, \end{aligned}$$

which proves the first assertion. Inequality (19) follows from (9) and the obvious identity

$$\theta_n(f) - \theta_{n-1}(f) = \frac{-1}{n(n-1)}\sum_{k=1}^{n-1} f(Y_k) + \frac{1}{n}f(Y_n) = \frac{1}{n}\left[f(Y_n) - \theta_{n-1}(f)\right] \;.$$

□

LEMMA 4.2. *Let $\alpha \in (0,1)$. Assume I1, I2a-b-c, I3a-b, and $\mathbb{E}[V(X_0)] < +\infty$. Then for any $\gamma, \gamma' \in (0,1)$ and any $\delta > \gamma$,*

$$n^{-\delta}\sum_{k=1}^n D_{V^\gamma}(\theta_k, \theta_{k-1})V^{\gamma'}(X_k) \xrightarrow{\mathbb{P}} 0 \;.$$

PROOF. By Lemma 4.1, we have

$$n^{-\delta}\sum_{k=1}^n D_{V^\gamma}(\theta_k, \theta_{k-1})V^{\gamma'}(X_k) \leq 2n^{-\delta}\sum_{k=1}^n \frac{1}{k}\left\{\theta_{k-1}(V^\gamma) + V^\gamma(Y_k)\right\}V^{\gamma'}(X_k) \;.$$

By I3-b, $\theta_k(V) \xrightarrow{\text{a.s.}} \theta_\star(V)$ thus implying that $\limsup_k\{\theta_k(V^\gamma) + k^{-\gamma}V^\gamma(Y_k)\} < \infty$, $\mathbb{P}$-a.s. . Therefore, the result holds if

$$\limsup_{n\to\infty} n^{-\delta}\sum_{k=1}^n k^{\gamma-1}\mathbb{E}[V^{\gamma'}(X_k)] = 0 \;.$$

Under the stated assumptions, Proposition 3.2 implies that $\sup_k \mathbb{E}\left[V(X_k)\right] < +\infty$ and this concludes the proof. □

LEMMA 4.3. *For any $\theta \in \Theta$, any measurable function $f : \mathsf{X} \to \mathbb{R}$ in $\mathcal{L}_{V^\alpha}$ and any $x, x' \in \mathsf{X}$ such that $\pi(x) \leq \pi(x')$*

$$\left|P_\theta f(x) - P_\theta f(x')\right| \leq \left|Pf(x) - Pf(x')\right| + \left|f(x) - f(x')\right| \\ + \sup_{\mathsf{X}} \pi \ |f|_{V^\alpha} \ \left|\pi^{-\beta}(x) - \pi^{-\beta}(x')\right| \ \left(V^\alpha(x') + \theta(V^\alpha)\right) \ .$$

PROOF. The proof is adapted from (Fort et al., 2010a, Lemma 5.1.); it is omitted for brevity. □

PROOF OF PROPOSITION 3.3. Let $\alpha \in (0, 1/2)$. By Proposition 3.2, **A2** and P[α] hold. By I3-b,

$$\limsup_n L_{\theta_n} < +\infty \ , \quad \mathbb{P} - \text{a.s.} \tag{20}$$

where L_θ is given by (1) with C_θ, ρ_θ defined by P[α].

We first check **A3-a**. Let $f \in \mathcal{N} \cap \mathcal{L}_{V^\alpha}$. By Lemma 5.1,

$$\left|P_{\theta_k}\Lambda_{\theta_k} f - P_{\theta_{k-1}}\Lambda_{\theta_{k-1}} f\right|_{V^\alpha} \leq 5 \ \left(L_{\theta_k} \vee L_{\theta_{k-1}}\right)^6 \pi_{\theta_k}(V^\alpha) D_{V^\alpha}(\theta_k, \theta_{k-1}) \ |f|_{V^\alpha} \ .$$

By Lemma 2.1, Proposition 3.2 and Assumptions I1, I2 and I3-b,

$$\limsup_{n\to\infty} \pi_{\theta_n}(V) \leq \tilde{b} \ (1 - \tilde{\lambda})^{-1} \limsup_{n\to\infty} \theta_n(V) < \infty \ , \ \mathbb{P} - \text{a.s.} \ . \tag{21}$$

Therefore, by (20) and (21), it suffices to prove that

$$n^{-1/2} \sum_{k=1}^n D_{V^\alpha}(\theta_k, \theta_{k-1}) V^\alpha(X_k) \xrightarrow{\mathbb{P}} 0 \ ,$$

which follows from Lemma 4.2. We now check **A3-b**. By Proposition 3.2, it holds

$$n^{-1/(2\alpha)} \sum_{k=1}^n L_{\theta_k}^{2/\alpha} P_{\theta_k} V(X_k) \leq n^{-1/(2\alpha)} \sum_{k=1}^n L_{\theta_k}^{2/\alpha} \left[V(X_k) + \tilde{b}\theta_k(V)\right] \ .$$

Under the stated assumptions, $\limsup_n \left[\theta_n(V) + L_{\theta_n}\right] < +\infty$ w.p.1. and by Proposition 3.2, $\sup_k \mathbb{E}\left[V(X_k)\right] < +\infty$. Since $2\alpha < 1$, this concludes the proof.

The proof of **A4** is in two steps: it is first proved that

$$\frac{1}{n} \sum_{k=0}^{n-1} F_{\theta_k}(X_k) - \frac{1}{n} \sum_{k=0}^{n-1} \int \pi_{\theta_k}(\mathrm{d}x) F_{\theta_k}(x) \xrightarrow{\mathbb{P}} 0 \ , \tag{22}$$

and then it is established that

$$\int \pi_{\theta_k}(\mathrm{d}x) F_{\theta_k}(x) \xrightarrow{\text{a.s.}} \int \pi_{\theta_\star}(\mathrm{d}x) F_{\theta_\star}(x) \; . \tag{23}$$

In order to prove (22), we check the conditions of Theorem 5.4 in Appendix 5 with $\gamma = 2\alpha$. First observe that $\Lambda_\theta f^2 \in \mathcal{L}_{V^{2\alpha}}$ (see (17)). We check conditions (i) to (vi) of Theorem 5.4.

(i) and (iii) follow from Proposition 3.2 and (20).

(ii) follows from Eq. (21).

(iv) follows from Lemma 4.2.

(v) under **A2**, we have by (17) and the Jensen's inequality

$$|F_\theta(x)| \leq 2 \, |f|_{V^\alpha}^2 \; L_\theta^4 \; P_\theta V^{2\alpha}(x) \leq 2 \, |f|_{V^\alpha}^2 \; L_\theta^4 \; \left| P_\theta V^{2\alpha} \right|_{V^{2\alpha}} V^{2\alpha}(x) \; .$$

Hence, $|F_\theta|_{V^{2\alpha}} \leq 2 \, |f|_{V^\alpha}^2 \; L_\theta^4 \left| P_\theta V^{2\alpha} \right|_{V^{2\alpha}}$. By I3-b, the drift inequality (14) and the Jensen's inequality, $\limsup_n \left| P_{\theta_n} V^{2\alpha} \right|_{V^{2\alpha}} < +\infty$ w.p.1. (20) concludes the proof.

(vi) Set $F_\theta(x) = G_\theta(x) - H_\theta(x)$ where $G_\theta(x) \stackrel{\text{def}}{=} P_\theta[\Lambda_\theta f]^2(x)$ and $H_\theta(x) \stackrel{\text{def}}{=} \left(P_\theta \Lambda_\theta f(x)\right)^2$, with

$$|\Lambda_\theta f|_{V^\alpha} + |\Lambda_{\theta'} f|_{V^\alpha} \leq M_{\theta,\theta'} \stackrel{\text{def}}{=} \left(L_\theta^2 + L_{\theta'}^2\right) |f|_{V^\alpha} \; .$$

By (20), $\limsup_{n\to\infty} M_{\theta_n,\theta_{n-1}} < \infty$, $\mathbb{P}$-a.s.

It holds

$$\begin{aligned}
&|G_\theta(x) - G_{\theta'}(x)| \\
&\leq \left| P_\theta \left(x, [\Lambda_\theta f]^2 - [\Lambda_{\theta'} f]^2\right) \right| + \left| \int \left\{ P_\theta(x, \mathrm{d}y) - P_{\theta'}(x, \mathrm{d}y) \right\} [\Lambda_{\theta'} f]^2(y) \right| \\
&\leq 2 M_{\theta,\theta'} \; P_\theta \left(x, |\Lambda_\theta f - \Lambda_{\theta'} f| \, V^\alpha\right) + M_{\theta,\theta'}^2 \; D_{V^{2\alpha}}(\theta,\theta') \; V^{2\alpha}(x) \\
&\leq 2 M_{\theta,\theta'} \; |\Lambda_\theta f - \Lambda_{\theta'} f|_{V^\alpha} \; \left| P_\theta V^{2\alpha} \right|_{V^{2\alpha}} V^{2\alpha}(x) + M_{\theta,\theta'}^2 \; D_{V^{2\alpha}}(\theta,\theta') \; V^{2\alpha}(x) \; .
\end{aligned}$$

By Lemma 5.1,

$$|f|_{V^\alpha}^{-1} \, |\Lambda_\theta f - \Lambda_{\theta'} f|_{V^\alpha} \leq 3 D_{V^\alpha}(\theta,\theta') \; \left(L_\theta \vee L_{\theta'}\right)^6 \pi_\theta(V^\alpha) \; .$$

Since w.p.1.:

$$\limsup_{n\to\infty} \left\{ \pi_{\theta_n}(V) + M_{\theta_n,\theta_{n-1}} + L_{\theta_n} + \left| P_{\theta_n} V^{2\alpha} \right|_{V^{2\alpha}} \right\} < \infty \; ,$$

it follows that $n^{-1} \sum_{k=1}^n V^{2\alpha}(X_k) \left| G_{\theta_k} - G_{\theta_{k-1}} \right|_{V^{2\alpha}}$ converges to zero in probability provided that

$$n^{-1} \sum_{k=1}^n \left[D_{V^{2\alpha}}(\theta_k, \theta_{k-1}) + D_{V^\alpha}(\theta_k, \theta_{k-1}) \right] V^{2\alpha}(X_k) \xrightarrow{\mathbb{P}} 0 \; ,$$

which follows from Lemma 4.2. Similarly, it holds

$$\begin{aligned}|H_\theta(x) - H_{\theta'}(x)| &\le |P_\theta \Lambda_\theta f(x) - P_{\theta'}\Lambda_{\theta'} f(x)|\,|P_\theta \Lambda_\theta f(x) + P_{\theta'}\Lambda_{\theta'} f(x)| \\ &\le M_{\theta,\theta'}\ |P_\theta \Lambda_\theta f(x) - P_{\theta'}\Lambda_{\theta'} f(x)|\,\{P_\theta V^\alpha(x) + P_{\theta'} V^\alpha(x)\}\ .\end{aligned}$$

Along the same lines as above, using Lemmas 4.2 and 5.1, we prove that $n^{-1}\sum_{k=1}^n V^{2\alpha}(X_k)\left|H_{\theta_k} - H_{\theta_{k-1}}\right|_{V^{2\alpha}}$ converges to 0 in probability.

We now consider the second step and prove (23). To that goal, we have to strengthen the conditions on f by assuming that f is continuous. For any $\theta \in \Theta$, $\int \pi_\theta(\mathrm{d}x) F_\theta(x) = \int \pi_\theta(\mathrm{d}x) H_\theta(x)$ with

$$H_\theta(x) \stackrel{\mathrm{def}}{=} \left(\Lambda_\theta f\right)^2(x) - \left(P_\theta \Lambda_\theta f\right)^2(x)\ . \tag{24}$$

We prove that there exists $\Omega_\star$ with $\mathbb{P}(\Omega_\star) = 1$ and for any $\omega \in \Omega_\star$,

(I) for any continuous bounded function h, $\lim_n \pi_{\theta_n(\omega)}(h) = \pi_{\theta_\star}(h)$,
(II) the set $\{H_{\theta_n(\omega)}, n \ge 0\}$ is equicontinuous,
(III) $\sup_n \pi_{\theta_n(\omega)}\left(|H_{\theta_n(\omega)}|^{1/(2\alpha)}\right) < +\infty$,
(IV) $\lim_n H_{\theta_n(\omega)}(x) = H_{\theta_\star}(x)$ for any $x \in \mathsf{X}$,
(V) $\pi_{\theta_\star}(|H_{\theta_\star}|) < +\infty$.

The proof is then concluded by application of Lemma 5.3.

Proof of (I). Under the conditions I1 and I2a-b-c, I3a-b and $\mathbb{E}[V(X_0)] < +\infty$, (Fort et al., 2010a, Proposition 3.3.) proves that this condition holds for any $\omega \in \Omega_1$ such that $\mathbb{P}(\Omega_1) = 1$. □

Proof of (II). Let C_θ, ρ_θ be given by P[α]. For any constants $C, v > 0$ and $\rho \in (0,1)$, set

$$\Theta_{C,\rho,v} \stackrel{\mathrm{def}}{=} \{\theta \in \Theta \text{ s.t. } C_\theta \le C, \rho_\theta \le \rho, \theta(V) \le v\}\ . \tag{25}$$

By (20) and I3-b,

$$\limsup_n C_{\theta_n} < +\infty\ , \mathbb{P}-\text{a.s.} \qquad \limsup_n \rho_{\theta_n} < 1\ , \mathbb{P}-\text{a.s.} \tag{26}$$

and $\limsup_n \theta_n(V) < +\infty$ w.p.1. Therefore, it is sufficient to prove that the set $\{H_\theta, \theta \in \Theta_{C,\rho,v}\}$ is equicontinuous.

Let $C, v > 0$ and $\rho \in (0,1)$ be fixed. Observe that by definition of L_θ (see (1)) and (17)

$$\sup_{\theta \in \Theta_{C,\rho,v}} |\Lambda_\theta f|_{V^\alpha} \le |f|_{V^\alpha}\ \left(C \vee (1-\rho)^{-1}\right)^2\ . \tag{27}$$

For any $x, x' \in \mathsf{X}$ and any $\theta \in \Theta_{C,\rho,v}$,

$$\begin{aligned}\left|H_\theta(x) - H_\theta(x')\right| &\leq \left|\Lambda_\theta f(x) + \Lambda_\theta f(x')\right| \left|\Lambda_\theta f(x) - \Lambda_\theta f(x')\right| \\ &+ \left|P_\theta \Lambda_\theta f(x) + P_\theta \Lambda_\theta f(x')\right| \left(\left|\Lambda_\theta f(x) - \Lambda_\theta f(x')\right| + \left|f(x) - f(x')\right|\right) ,\end{aligned}$$

where we have used $P_\theta \Lambda_\theta f(x) - P_\theta \Lambda_\theta f(x') = (\Lambda_\theta - \mathrm{I})[f(x) - f(x')]$. By (14) and (27), for any $\theta \in \Theta_{C,\rho,v}$,

$$|P_\theta \Lambda_\theta f(x)| \leq |f|_{V^\alpha} \ \left(C \vee (1-\rho)^{-1}\right)^2 \left(V(x) + \tilde{b} v\right) . \tag{28}$$

Therefore, since f and V are continuous, it suffices to prove that the set $\{\Lambda_\theta f, \theta \in \Theta_{C,\rho,v}\}$ is equicontinuous. Lemma 4.3 and I2-d imply that the set $\{P_\theta f, \theta \in \Theta_{C,\rho,v}\}$ is equicontinuous. Repeated applications of this Lemma shows that for any $\ell \geq 1$, the set $\{P^\ell_\theta f, \theta \in \Theta_{C,\rho,v}\}$ is equicontinuous. By Proposition 3.2, we have for any $\theta \in \Theta_{C,\rho,v}$,

$$\begin{aligned}\left|\Lambda_\theta f(x) - \Lambda_\theta f(x')\right| \leq \sum_{k=0}^{n-1} &\left|P^k_\theta f(x) - P^k_\theta f(x')\right| \\ &+ 2\rho^n \, C \, (1-\rho)^{-1} \ \left(V(x) + V(x')\right) .\end{aligned}$$

Then, the set $\{\Lambda_\theta f, \theta \in \Theta_{C,\rho,v}\}$ is equicontinuous. □

Proof of (III). By (20) and (21), there exists Ω_3 such that $\mathbb{P}(\Omega_3) = 1$ and for any $\omega \in \Omega_3$, (III) holds if, for any constants $C, v > 0$ and $\rho \in (0,1)$,

$$\sup_{\theta \in \Theta_{C,\rho,v}} |H_\theta|^{1/2\alpha}_{V^{2\alpha}} \ \pi_\theta(V) < +\infty ,$$

where $\Theta_{C,\rho,v}$ is defined by (25). The bound on H_θ follows from (27) and (28). By Lemma 2.1 and Proposition 3.2, $\sup_{\theta \in \Theta_{C,\rho,v}} \pi_\theta(V) \leq (1-\tilde{\lambda})^{-1} \ \tilde{b} \ v$. □

Proof of (IV). We first prove that for any $x \in \mathsf{X}$, $\lim_n \Lambda_{\theta_n} f(x) \xrightarrow{\text{a.s.}} \Lambda_{\theta_\star} f(x)$. By Proposition 3.2, for any $\ell \geq 1$,

$$\begin{aligned}|\Lambda_{\theta_n} f(x) - \Lambda_{\theta_\star} f(x)| \leq C_{\theta_n} \rho^\ell_{\theta_n} V(x) + C_{\theta_\star} \rho^\ell_{\theta_\star} V(x) + |\pi_{\theta_n}(f) - \pi_\star(f)| \\ + \sum_{k=0}^{\ell-1} \left|P^k_{\theta_n} f(x) - P^k_{\theta_\star} f(x)\right| .\end{aligned} \tag{29}$$

From (Fort et al., 2010a, Proposition 3.3.), $\pi_{\theta_n}(f) - \pi_{\theta_\star}(f) \xrightarrow{\text{a.s.}} 0$ since f is continuous. In addition, following the same lines as in the proof of (Fort et al., 2010a, Proposition 3.3, Lemma 4.4.), it holds $P^k_{\theta_n} f(x) - P^k_{\theta_\star} f(x) \xrightarrow{\text{a.s.}} 0$ for any k. Therefore, by (26), (29) shows that $\lim_n \Lambda_{\theta_n} f(x) \xrightarrow{\text{a.s.}} \Lambda_{\theta_\star} f(x)$.

It remains to prove that $\lim_n P_{\theta_n} \Lambda_{\theta_n} f(x) \xrightarrow{\text{a.s.}} P_{\theta_\star} \Lambda_{\theta_\star} f(x)$. This is a consequence of the above discussion and the equality

$$P_{\theta_n} \Lambda_{\theta_n} f(x) - P_{\theta_\star} \Lambda_{\theta_\star} f(x) = \Lambda_{\theta_n} f(x) - \Lambda_{\theta_\star} f(x) + \pi_{\theta_n}(f) - \pi_{\theta_\star}(f) \; ,$$

which follows from (5).

Combining the two results above, for any $x \in \mathsf{X}$, $H_{\theta_n}(x) \xrightarrow{\text{a.s.}} H_{\theta_\star}(x)$ as $n \to +\infty$. Since X is Polish, there exists a countable dense subset $\mathcal{D}$ of X and a set Ω_4 with $\mathbb{P}(\Omega_4) = 1$ such that for any $\bar{x} \in \mathcal{D}$ and any $\omega \in \Omega_4$,

$$\lim_n H_{\theta_{n(\omega)}}(\bar{x}) = H_{\theta_\star}(\bar{x}) \; .$$

The proof is concluded by the inequality

$$\begin{aligned} \left| H_{\theta_{n(\omega)}}(x) - H_{\theta_\star}(x) \right| &\leq \left| H_{\theta_{n(\omega)}}(x) - H_{\theta_{n(\omega)}}(\bar{x}) \right| \\ &\quad + \left| H_{\theta_{n(\omega)}}(\bar{x}) - H_{\theta_\star}(\bar{x}) \right| + \left| H_{\theta_\star}(\bar{x}) - H_{\theta_\star}(x) \right| \; , \end{aligned}$$

the continuity of $H_{\theta_\star}$ and (II). □

Proof (V). Since $H_{\theta_\star} \in \mathcal{L}_{V^{2\alpha}}$, this is a consequence of Lemma 2.1 and Assumption I3-a. □

□

4.2.3. *Proof of Theorem 3.4.* We check the conditions of Theorem 2.3. **A2** to **A5** hold (see Propositions 3.2 and 3.3) and we now prove **A6**. We first check condition **A6**-a. For any function $f \in \mathcal{L}_{V^\alpha} \cap \mathcal{N}$, define

(30)
$$G_f(z) \overset{\text{def}}{=} \epsilon \iint \left(\delta_z(\mathrm{d}z') - \theta_\star(\mathrm{d}z') \right) \pi_{\theta_\star}(\mathrm{d}x) r(x, z') \left(\Lambda_{\theta_\star} f(z') - \Lambda_{\theta_\star} f(x) \right) \; .$$

Let $f \in \mathcal{L}_{V^\alpha} \cap \mathcal{N}$; note that $G_f \in \mathcal{L}_{V^\alpha}$. Recall that by Eq. (11), for any θ such that $\theta(V^\alpha) < +\infty$,

$$P_\theta f(x) - P_{\theta_\star} f(x) = \epsilon \int \left[\theta(\mathrm{d}y) - \theta_\star(\mathrm{d}y) \right] r(x, y) \left(f(y) - f(x) \right) \; . \tag{31}$$

Then, using (30),

$$\begin{aligned} &\pi_{\theta_\star} \left(P_{\theta_k} - P_{\theta_\star} \right) \Lambda_{\theta_\star} f \\ &= \epsilon \iint \pi_{\theta_\star}(\mathrm{d}x) \left[\theta_k(\mathrm{d}z) - \theta_\star(\mathrm{d}z) \right] r(x, z) \left[\Lambda_{\theta_\star} f(z) - \Lambda_{\theta_\star} f(x) \right] = \theta_k(G_f) \; . \end{aligned}$$

Therefore,

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\pi_{\theta_\star}\left(P_{\theta_k}-P_{\theta_\star}\right)\Lambda_{\theta_\star}f &= \frac{1}{n}\sum_{k=1}^{n}\frac{n}{k}\frac{1}{\sqrt{n}}\sum_{j=1}^{k}G_f(Y_j)\\ &= \int_0^1 t^{-1}S_n(t)\mathrm{d}t+\sum_{k=1}^{n-1}\int_{k/n}^{(k+1)/n}\left(\frac{n}{k}-\frac{1}{t}\right)S_n(t)\mathrm{d}t+\frac{1}{n}S_n(1)\;,\end{aligned}$$

with $S_n(t) \stackrel{\text{def}}{=} n^{-1/2}\sum_{j=1}^{[nt]} G_f(Y_j)$. Note that

$$\mathbb{E}\left[\left|\sum_{k=1}^{n-1}\int_{k/n}^{(k+1)/n}\left(\frac{n}{k}-\frac{1}{t}\right)S_n(t)\mathrm{d}t\right|\right] \leq \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\frac{1}{k+1}\frac{1}{k}\sum_{j=1}^{k}\mathbb{E}\left[|G_f(Y_j)|\right]\;.$$

Since $G_f \in \mathcal{L}_{V^\alpha}$, I3-a implies that $\sup_{k\geq 0}\mathbb{E}[|G_f|(Y_k)] < \infty$. Therefore,

$$\sum_{k=1}^{n-1}\int_{k/n}^{(k+1)/n}\left(\frac{n}{k}-\frac{1}{t}\right)S_n(t)\mathrm{d}t+\frac{1}{n}S_n(1) \xrightarrow{\mathbb{P}} 0\;.$$

Using I3-c and the Continuous mapping Theorem ((van der Vaart and Wellner, 1996, Theorem 1.3.6)), we obtain

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\pi_{\theta_\star}\left(P_{\theta_k}-P_{\theta_\star}\right)\Lambda_{\theta_\star}f \xrightarrow{\mathcal{D}} \tilde{\gamma}^2(f)\int_0^1 t^{-1}B_t\mathrm{d}t\;.$$

Since $\int_0^1 t^{-1}B_t\mathrm{d}t = \int_0^1\log(t)\mathrm{d}B_t$, $\int_0^1 t^{-1}B_t\mathrm{d}t$ is a Gaussian random variable with zero mean and variance $\int_0^1\log^2(t)\mathrm{d}t = 2$.

We now check condition **A6**-b. Note that

$$n^{-1/2}\sum_{k=1}^{n}\pi_{\theta_k}\left(P_{\theta_k}-P_{\theta_\star}\right)\Lambda_{\theta_\star}\left(P_{\theta_k}-P_{\theta_\star}\right)\Lambda_{\theta_\star}f = n^{-1/2}\sum_{k=1}^{n}\pi_{\theta_k}(G^f_{\theta_k})\;,$$

where

$$G^f_\theta(x) \stackrel{\text{def}}{=} \left(P_\theta-P_{\theta_\star}\right)\Lambda_{\theta_\star}\left(P_\theta-P_{\theta_\star}\right)\Lambda_{\theta_\star}f(x)\;. \tag{32}$$

We write for any $x \in \mathsf{X}$ and any $\ell_k \in \mathbb{N}$,

$$\pi_{\theta_k}(G^f_{\theta_k}) = \left(\pi_{\theta_k}-P^{\ell_k}_{\theta_k}\right)G^f_{\theta_k}(x)+\left(P^{\ell_k}_{\theta_k}-P^{\ell_k}_{\theta_\star}\right)G^f_{\theta_k}(x)+P^{\ell_k}_{\theta_\star}G^f_{\theta_k}(x)\;.$$

By Proposition 3.2, P$[\alpha]$ holds and there exist C_θ, ρ_θ such that $\|P_\theta^n - \pi_\theta\|_{V^\alpha} \leq C_\theta \rho_\theta^n$. Furthermore, Lemma 5.2 and I3b imply that $\limsup_n C_{\theta_n} < +\infty$ w.p.1. and there exists a constant $\rho \in (0,1)$ such that $\limsup_n \rho_{\theta_n} \leq \rho$, w.p. 1. Set $\ell_k \stackrel{\text{def}}{=} \lfloor \ell \ln k \rfloor$ with ℓ such that $1/2 + \ell \ln \rho < 0$. Let $x \in \mathsf{X}$ be fixed.

By Lemma 4.4 and I3-b, there exists an almost surely finite random variable C_1 s.t.

$$\left| \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \left(\pi_{\theta_k} - P_{\theta_k}^{\ell_k} \right) G_{\theta_k}^f(x) \right| \leq C_1 V^\alpha(x) n^{-1/2} \sum_{k=1}^{n} \rho^{\ell_k} \ .$$

Since $n^{-1/2} \sum_{k=1}^n \rho^{\ell_k} \leq \rho^{-1} n^{-1/2} \sum_{k=1}^n k^{\ell \ln \rho} \to_{n \to \infty} 0$, it holds

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \left(\pi_{\theta_k} - P_{\theta_k}^{\ell_k} \right) G_{\theta_k}^f(x) \xrightarrow{\text{a.s.}} 0 \ .$$

By Lemma 4.6, there exist some positive constants $C_2, \kappa_\star, a$ such that

$$\mathbb{E}\left[\left(\sum_{k=1}^{n} \{ P_{\theta_k}^{\ell_k} - P_{\theta_\star}^{\ell_k} \} G_{\theta_k}^f(x) \right)^2 \right]^{1/2} \leq C_2 \, |f|_{V^\alpha} \, V^\alpha(x) \sum_{k=1}^{n} \frac{1}{k} \sum_{t=1}^{\ell_k - 1} \left(\frac{\kappa_\star \ell_k}{k^{1/(2a)}} \right)^{at} \ .$$

Since $\lim_k \ell_k^a / k^{1/2} = 0$, there exists $k_\star$ such that for $k \geq k_\star$, $(\kappa_\star \ell_k)^a / k^{1/2} \leq 1/2$. Then,

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{1}{k} \sum_{t=1}^{\ell_k} \left(\frac{\kappa_\star \ell_k}{k^{1/(2a)}} \right)^{at} \leq \frac{1}{\sqrt{n}} \sum_{k=1}^{k_\star} \frac{1}{k} \sum_{t=1}^{\lceil \ell \ln k \rceil} \left(\frac{\kappa_\star \ell_k}{k^{1/(2a)}} \right)^{at} + \frac{2}{\sqrt{n}} \sum_{k=k_\star+1}^{n} \frac{1}{k} \ .$$

The RHS tends to zero when $n \to +\infty$, which proves that $n^{-1/2} \sum_{k=1}^n \{ P_{\theta_k}^{\ell_k} - P_{\theta_\star}^{\ell_k} \} G_{\theta_k}^f(x) \xrightarrow{\mathbb{P}} 0$.

Finally, by Lemma 4.7, there exists a constant C_3 such that

$$\mathbb{E}\left[\left(\frac{1}{\sqrt{n}} \sum_{k=1}^{n} P_{\theta_\star}^{\ell_k} G_{\theta_k}^f(x) \right)^2 \right]^{1/2} \leq C_3 V^\alpha(x) \, \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{\ell_k^\alpha}{k} \to_{n \to \infty} 0 \ ,$$

thus implying that $n^{-1/2} \sum_{k=1}^n P_{\theta_\star}^{\ell_k} G_{\theta_k}^f(x) \xrightarrow{\mathbb{P}} 0$.

LEMMA 4.4. *Assume I1 and I2a-b-c. Let $\alpha \in (0, 1/2)$. For any $f \in \mathcal{L}_{V^\alpha}$ and $\theta \in \Theta$,*

$$G_\theta^f(x) = \int (\theta - \theta_\star)^{\otimes 2}(\mathrm{d}z_{1:2}) \ F^{(0)}(x, z_1, z_2) \ ,$$

where G_θ^f is defined by (32); and there exists a constant C such that for any $x \in \mathsf{X}$,

$$\left| F^{(0)}(x, z_1, z_2) \right| \le C \, |f|_{V^\alpha} \ V^{\alpha \wedge (\beta/\tau)}(x) \ \left(V^\alpha(z_1) + V^\alpha(z_2) \right) \ .$$

In addition, there exists some constant C' such that for any $\ell \in \mathbb{N}$, any $\theta \in \Theta$ and any $f \in \mathcal{L}_{V^\alpha}$,

$$\left| \left(\pi_\theta - P_\theta^\ell \right) G_\theta^f \right|_{V^\alpha} \le C' \, |f|_{V^\alpha} \ \left\| P_\theta^\ell - \pi_\theta \right\|_{V^\alpha} \ \theta(V^\alpha) \ .$$

Proof. Set $\gamma \stackrel{\text{def}}{=} \alpha \wedge (\beta/\tau)$. Throughout this proof, let L_θ be the constant given by P[γ]. We have

$$F^{(0)}(x, z_1, z_2) \stackrel{\text{def}}{=} \epsilon^2 r(x, z_2) \left[\int \Lambda_{\theta_\star}(z_2, \mathrm{d}y) r(y, z_1) \left(\Lambda_{\theta_\star} f(z_1) - \Lambda_{\theta_\star} f(y) \right) \right.$$
$$\left. - \int \Lambda_{\theta_\star}(x, \mathrm{d}y) r(y, z_1) \left(\Lambda_{\theta_\star} f(z_1) - \Lambda_{\theta_\star} f(y) \right) \right] \ .$$

Note that $|r(\cdot, z_1)|_{V^\gamma} \le 1$ for any z_1 so that by (17),

$$\left| \int \Lambda_{\theta_\star}(z_2, \mathrm{d}y) r(y, z_1) \Lambda_{\theta_\star} f(z_1) \right| \le L_{\theta_\star}^4 \, |f|_{V^\alpha} \, V^\alpha(z_1) \ V^\gamma(z_2) \ .$$

In addition, since $\gamma - \beta/\tau \le 0$, we have by definition of the acceptance ratio r (see (10))

$$r(x, z_2) V^\gamma(z_2) \le V^\gamma(x) \ .$$

Then, there exists a constant C such that

$$\epsilon^2 r(x, z_2) \left| \int \Lambda_{\theta_\star}(z_2, \mathrm{d}y) r(y, z_1) \Lambda_{\theta_\star} f(z_1) \right| \le C \, |f|_{V^\alpha} \, V^\alpha(z_1) \ V^\gamma(x) \ .$$

Similar upper bounds can be obtained for the three remaining terms in $F^{(0)}$, thus showing the upper bounds on $F^{(0)}$.

In addition, by P[γ]

$$\left| \left(\pi_\theta - P_\theta^\ell \right) G_\theta^f f(x) \right|_{V^\alpha} \le \left\| \pi_\theta - P_\theta^\ell \right\|_{V^\alpha} \ \left| G_\theta^f f \right|_{V^\alpha} V^\alpha(x) \ .$$

The proof is concluded upon noting that $|G_\theta^f(x)| \le C \, |f|_{V^\alpha} \, \theta(V^\alpha)$. □

LEMMA 4.5. *Assume I1 and I2a-b-c. Let $\alpha \in (0, 1/2)$. There exist some constants $C, \kappa_\star$ and $\rho_\star \in (0,1)$ such that for any $t \geq 1$, any integers $u_1, \cdots, u_t$ and any $f \in \mathcal{L}_{V^\alpha}$,*

$$
\begin{aligned}
&(P_\theta - P_{\theta_\star}) \left(P_{\theta_\star}^{u_t} - \pi_{\theta_\star}\right) \cdots (P_\theta - P_{\theta_\star}) \left(P_{\theta_\star}^{u_1} - \pi_{\theta_\star}\right) G_\theta^f(x) \\
&\qquad = \int \cdots \int (\theta - \theta_\star)^{\otimes(t+2)} \left(\mathrm{d}z_{1:t+2}\right) F_{u_{1:t}}^{(t)}(x, z_1, \cdots, z_{t+2})
\end{aligned}
$$

where G_θ^f is defined in (32), and

$$
\left|F_{u_{1:t}}^{(t)}(x, z_1, \cdots, z_{t+2})\right| \leq C \left|f\right|_{V^\alpha} \kappa_\star^t \; \rho_\star^{\sum_{j=1}^t u_j} V^{\alpha \wedge (\beta/\tau)}(x) \sum_{j=1}^{t+2} V^\alpha(z_j) \; . \tag{33}
$$

PROOF. By repeated applications of Eq. (31), it can be proved that the functions $F_{u_{1:t}}^{(t)}$ are recursively defined as follows

$$
\begin{aligned}
F_{u_{1:t}}^{(t)}(x, z_1, \cdots, z_{t+2}) &\stackrel{\text{def}}{=} \epsilon r(x, z_{t+2}) \times \\
&\int \left(P_{\theta_\star}^{u_t}(z_{t+2}, \mathrm{d}y) - P_{\theta_\star}^{u_t}(x, \mathrm{d}y)\right) F_{u_{1:t-1}}^{(t-1)}(y, z_1, \cdots, z_{t+1}) \; ,
\end{aligned} \tag{34}
$$

where $F_{u_{1:0}}^{(0)} = F^{(0)}$ and $F^{(0)}$ is given by Lemma 4.4.

The proof of the upper bound is by induction. The property holds for $t = 1$. Assume it holds for $t \geq 2$. Set $\gamma \stackrel{\text{def}}{=} \alpha \wedge (\beta/\tau)$; by Proposition 3.2 and the property P$[\gamma]$, there exist some constants $C_\star$ and $\rho_\star \in (0,1)$ such that $\left\|P_{\theta_\star}^\ell - \pi_{\theta_\star}\right\|_{V^\gamma} \leq C_{\theta_\star} \rho_{\theta_\star}^\ell$. Then,

$$
\begin{aligned}
\left|F_{u_{1:t}}^{(t)}(x, z_{1:t+2})\right| &\leq C \left|f\right|_{V^\alpha} \kappa_\star^{t-1} \rho_{\theta_\star}^{\sum_{j=1}^{t-1} u_j} \left(\sum_{j=1}^{t+1} V^\alpha(z_j)\right) \\
&\qquad \times r(x, z_{t+2}) \left[\left\|P_{\theta_\star}^{u_t} - \pi_{\theta_\star}\right\|_{V^\gamma} V^\gamma(z_{t+2}) + \left\|P_{\theta_\star}^{u_t} - \pi_{\theta_\star}\right\|_{V^\gamma} V^\gamma(x)\right] \\
&\leq C \; \left|f\right|_{V^\alpha} \; \kappa_\star^{t-1} \; \epsilon \; C_{\theta_\star} \; \rho_{\theta_\star}^{\sum_{j=1}^t u_j} \; r(x, z_{t+2}) \; \left\{V^\gamma(z_{t+2}) + V^\gamma(x)\right\} \; .
\end{aligned}
$$

Since $\gamma \leq \beta/\tau$, $r(x, z_{t+2}) V^\gamma(z_{t+2}) \leq V^\gamma(x)$ thus showing (33) with $\kappa_\star = 2 C_{\theta_\star} \epsilon$. □

LEMMA 4.6. *Assume I1, I2a-b-c and I3. Let $\alpha \in (0, 1/2)$. There exist positive constants C, κ, a such that for any $f \in \mathcal{L}_{V^\alpha}$, any $k, \ell \geq 1$ and any $x \in \mathsf{X}$,*

$$
\mathbb{E}\left[\left(\left\{P_{\theta_k}^\ell - P_{\theta_\star}^\ell\right\} G_{\theta_k}^f(x)\right)^2\right]^{1/2} \leq C \left|f\right|_{V^\alpha} \frac{V^\alpha(x)}{k} \sum_{t=1}^{\ell-1} \left(t \kappa k^{-1/(2a)}\right)^{at} \; ,
$$

where G^f_θ is given by (32).

PROOF. For any $g \in \mathcal{L}_{V^\alpha}$, $k, \ell \geq 1$ and $x \in \mathsf{X}$,

$$\begin{aligned}
&P^\ell_{\theta_k} g(x) - P^\ell_{\theta_\star} g(x) \\
&= \sum_{t=1}^{\ell-1} \sum_{u_{1:t} \in \mathcal{U}_t} P_{\theta_\star}^{\ell-t-\sum_{j=1}^t u_j} \left(P_{\theta_k} - P_{\theta_\star}\right) P_{\theta_\star}^{u_t} \cdots \left(P_{\theta_k} - P_{\theta_\star}\right) P_{\theta_\star}^{u_1} g(x) \;, \\
&= \sum_{t=1}^{\ell-1} \sum_{u_{1:t} \in \mathcal{U}_t} P_{\theta_\star}^{\ell-t-\sum_{j=1}^t u_j} \left(P_{\theta_k} - P_{\theta_\star}\right) \left(P_{\theta_\star}^{u_t} - \pi_{\theta_\star}\right) \\
&\qquad\qquad \times \cdots \left(P_{\theta_k} - P_{\theta_\star}\right) \left(P_{\theta_\star}^{u_1} - \pi_{\theta_\star}\right) g(x) \;,
\end{aligned}$$

where $\mathcal{U}_t = \{u_{1:t}, u_j \in \mathbb{N}, \sum_{j=1}^t u_j \leq \ell - t\}$. Fix $t \in \{1, \cdots, \ell-1\}$ and $u_{1:t} \in \mathcal{U}_t$. Then by Lemma 4.5,

$$\begin{aligned}
&P_{\theta_\star}^{\ell-t-\sum_{j=1}^t u_j} \left(P_{\theta_k} - P_{\theta_\star}\right) \left(P_{\theta_\star}^{u_t} - \pi_{\theta_\star}\right) \cdots \left(P_{\theta_k} - P_{\theta_\star}\right) \left(P_{\theta_\star}^{u_1} - \pi_{\theta_\star}\right) G^f_{\theta_k}(x) \\
&= \int (\theta_k - \theta_\star)^{\otimes(t+2)} \left(\mathrm{d}z_{1:t+2}\right) \int P_{\theta_\star}^{\ell-t-\sum_{j=1}^t u_j}(x, \mathrm{d}y) F^{(t)}_{u_{1:t}}(y, z_1, \cdots, z_{t+2}) \;.
\end{aligned}$$

Assumptions I3-b and I3-d and Lemma 4.5 show that there exist constants $C, \kappa_\star, \rho_\star \in (0,1)$ such that

$$\begin{aligned}
&\left\| \int (\theta_k - \theta_\star)^{\otimes(t+2)} \left(\mathrm{d}z_{1:t+2}\right) \int P_{\theta_\star}^{\ell-t-\sum_{j=1}^t u_j}(x, \mathrm{d}y) F^{(t)}_{u_{1:t}}(y, z_1, \cdots, z_{t+2}) \right\|_2 \\
&\qquad \leq \frac{C}{k^{1+t/2}} A_t \left|f\right|_{V^\alpha} \kappa_\star^t \; \rho_\star^{\sum_{j=1}^t u_j} \; P_{\theta_\star}^{\ell-t-\sum_{j=1}^t u_j} V^\alpha(x) \;.
\end{aligned}$$

Finally, Proposition 3.2 implies that $\sup_{j\geq 0} \left|P^j_{\theta_\star} V^\alpha\right|_{V^\alpha} < +\infty$. By combining these results, we have for some constant C

$$\left\| P^\ell_{\theta_k} G^f_{\theta_k}(x) - P^\ell_{\theta_\star} G^f_{\theta_k}(x) \right\|_2 \leq C k^{-1} \left|f\right|_{V^\alpha} V^\alpha(x) \sum_{t=1}^{\ell-1} A_t \kappa_\star^t \, k^{-t/2} \sum_{u_{1:t} \in \mathcal{U}_t} \rho_\star^{\sum_{j=1}^t u_j} \;.$$

Note that $\sum_{u_{1:t} \in \mathcal{U}_t} \rho_\star^{\sum_{j=1}^t u_j} \leq (1-\rho_\star)^{-t}$. Furthermore, there exists $a > 0$ such that $A_t \leq t^{at}$. Therefore,

$$\begin{aligned}
&\left\| P^\ell_{\theta_k} G^f_{\theta_k}(x) - P^\ell_{\theta_\star} G^f_{\theta_k}(x) \right\|_2 \\
&\qquad \leq C k^{-1} \left|f\right|_{V^\alpha} V^\alpha(x) \sum_{t=1}^{\ell-1} \left(t \kappa^{1/a} (1-\rho_\star)^{-1/a} \, k^{-1/(2a)} \right)^{at} \;.
\end{aligned}$$

This concludes the proof. $\square$

LEMMA 4.7. *Assume I1, I2a-b-c and I3. Let $\alpha \in (0, 1/2)$ and $f \in \mathcal{L}_{V^\alpha}$. Then, there exists a constant C such that for any $k, \ell \geq 1$ and any $x \in \mathsf{X}$,*

$$\mathbb{E}\left[\left(P^\ell_{\theta_\star} G^f_{\theta_k}(x)\right)^2\right]^{1/2} \leq C \; \ell^\alpha \; |f|_{V^\alpha} \; k^{-1} V^\alpha(x) \; .$$

PROOF. We have

$$P^\ell_{\theta_\star} G^f_{\theta_k}(x) = \iint (\theta_k - \theta_\star)^{\otimes 2} (\mathrm{d}z_{1:2}) H_\ell(x, z_1, z_2) \; ,$$

with $H_\ell(x, z_1, z_2) \stackrel{\mathrm{def}}{=} P^\ell_{\theta_\star}(x, F^{(0)}(\cdot, z_1, z_2))$ where $F^{(0)}$ is given by Lemma 4.4. Lemma 4.4 also implies that there exists a constant C such that

$$|H_\ell(x, z_1, z_2)| \leq C \; |f|_{V^\alpha} \left(V^\alpha(z_1) + V^\alpha(z_2)\right) \; P^\ell_{\theta_\star} V^\alpha(x) \; . \tag{35}$$

By I3, the variance of $P^\ell_{\theta_\star} G^f_{\theta_k}(x)$ is upper bounded by

$$C \, |f|^2_{V^\alpha} \, (P^\ell_{\theta_\star} V^\alpha(x))^2 k^{-2} \; .$$

The proof is concluded by application of the drift inequality (14) and I3-a. $\square$

5. Appendix.

5.1. *Technical lemmas.* The following lemma is (slightly) adapted from (Fort et al., 2010a, Lemma 4.2.)

LEMMA 5.1. *Assume **A2**. For any $f \in \mathcal{L}_{V^\alpha}$ and $\theta, \theta' \in \Theta$,*

$$\begin{aligned}
&\|\pi_\theta - \pi_{\theta'}\|_{V^\alpha} \leq 2(L_{\theta'} \vee L_\theta)^4 \pi_\theta(V^\alpha) \; D_{V^\alpha}(\theta, \theta') \; , \\
&\|\Lambda_\theta - \Lambda_{\theta'}\|_{V^\alpha} \leq 3 \; \left(L_\theta \vee L_{\theta'}\right)^6 \pi_\theta(V^\alpha) D_{V^\alpha}(\theta, \theta') \\
&\|P_\theta \Lambda_\theta - P_{\theta'} \Lambda_{\theta'}\|_{V^\alpha} \leq 5 \; \left(L_\theta \vee L_{\theta'}\right)^6 \pi_\theta(V^\alpha) D_{V^\alpha}(\theta, \theta') \; .
\end{aligned}$$

where L_θ and Λ_θ are given by (1) *and (4).*

The following lemma can be obtained from Roberts and Rosenthal (2004), Fort and Moulines (2003), Douc et al. (2004) or Baxendale (2005) (see also the proof of (Saksman and Vihola, 2010, Lemma 3) for a similar result).

LEMMA 5.2. *Let $\{P_\theta, \theta \in \Theta\}$ be a family of phi-irreducible and aperiodic Markov kernels. Assume that there exist a function $V : \mathsf{X} \to [1, +\infty)$, and for any $\theta \in \Theta$ there exist some constants $b_\theta < \infty$, $\delta_\theta \in (0,1)$, $\lambda_\theta \in (0,1)$ and a probability measure ν_θ on X such that*

$$P_\theta V \leq \lambda_\theta V + b_\theta \;,$$

$$P_\theta(x, \cdot) \geq \delta_\theta \; \nu_\theta(\cdot) \; \mathbb{1}_{\{V \leq c_\theta\}}(x) \qquad c_\theta \stackrel{\text{def}}{=} 2b_\theta(1-\lambda_\theta)^{-1} - 1 \;.$$

Then there exists $\gamma > 0$ and for any θ, there exist some finite constants C_θ and $\rho_\theta \in (0,1)$ such that

$$\|P_\theta^n(x, \cdot) - \pi_\theta\|_V \leq C_\theta \; \rho_\theta^n \; V(x)$$

and

$$C_\theta \vee (1-\rho_\theta)^{-1} \leq C \left\{ b_\theta \vee \delta_\theta^{-1} \vee (1-\lambda_\theta)^{-1} \right\}^\gamma \;.$$

Lemma 5.3 is proved in (Fort et al., 2010b, Section 4).

LEMMA 5.3. *Let X be a Polish space endowed with its Borel σ-field $\mathcal{X}$. Let μ and $(mu_n)_{n\in\mathbb{N}}$ be probability distributions on $(\mathsf{X}, \mathcal{X})$. Let $(h_n)_{n\in\mathbb{N}}$ be an equicontinuous family of functions from X to $\mathbb{R}$. Assume*

(i) the sequence $(\mu_n)_{n\in\mathbb{N}}$ converges weakly to μ,
(ii) for any $x \in \mathsf{X}$, $\lim_{n\to\infty} h_n(x)$ exists, and there exists $\gamma > 1$ such that $\sup_n \mu_n(|h_n|^\gamma) + \mu(|\lim_n h_n|) < +\infty$.

Then, $\mu_n(h_n) \to \mu(\lim_n h_n)$.

5.2. *Weak law of large numbers for adaptive and interacting MCMC algorithms.* The proof of the theorem below is along the same lines as the proof of (Fort et al., 2010a, Theorem 2.7), which addresses the strong law of large numbers and details are omitted. Note that in this generalization, we relax the condition $\sup_\theta |F(\cdot, \theta)|_V < +\infty$ of Fort et al. (2010a). The proof is provided in the supplementary paper (Fort et al., 2011).

THEOREM 5.4. *Assume **A1**, **A2** and let $\gamma \in (0,1)$. Assume that*

(i) $\limsup_{n\to\infty} L_{\theta_n} < \infty$, $\mathbb{P}$-a.s. where L_θ is defined in Lemma 2.1 applied with the closed interval $[\gamma, 1]$.
(ii) $\limsup_{n\to\infty} \pi_{\theta_n}(V^\gamma) < \infty$, $\mathbb{P}$-a.s. .
(iii) $\sup_{k\geq 1} \mathbb{E}[V(X_k)] < \infty$.
(iv) $n^{-1} \sum_{k=1}^n D_{V^\gamma}(\theta_k, \theta_{k-1}) V^\gamma(X_k) \xrightarrow{\mathbb{P}} 0$.

Let $F : \mathsf{X} \times \Theta \to \mathbb{R}$ be a measurable function s.t.

(v) $\limsup_{n\to\infty} |F_{\theta_n}|_{V^\gamma} < +\infty$.

(vi) $n^{-1} \sum_{k=1}^{n-1} \left|F_{\theta_k} - F_{\theta_{k-1}}\right|_{V^\gamma} V^\gamma(X_k) \xrightarrow{\mathbb{P}} 0$.

Then,

$$\frac{1}{n}\sum_{k=0}^{n-1} F_{\theta_k}(X_k) - \frac{1}{n}\sum_{k=0}^{n-1}\int \pi_{\theta_k}(\mathrm{d}x) F_{\theta_k}(x) \xrightarrow{\mathbb{P}} 0 \ .$$

References.

Andrieu, C., A. Jasra, A. Doucet, and P. Del Moral (2007a). Convergence of the equi-energy sampler. In *Conference Oxford sur les méthodes de Monte Carlo séquentielles*, Volume 19 of *ESAIM Proc.*, pp. 1–5.

Andrieu, C., A. Jasra, A. Doucet, and P. Del Moral (2007b). Non-linear Markov chain Monte Carlo. In *Conference Oxford sur les méthodes de Monte Carlo séquentielles*, Volume 19 of *ESAIM Proc.*, pp. 79–84.

Andrieu, C., A. Jasra, A. Doucet, and P. Del Moral (2008). A note on convergence of the equi-energy sampler. *Stoch. Anal. Appl. 26*(2), 298–312.

Andrieu, C., A. Jasra, A. Doucet, and P. Del Moral (2011). On non-linear Markov chain Monte Carlo via self-interacting approximations. *Bernoulli (to appear)*.

Andrieu, C. and E. Moulines (2006). On the ergodicity property of some adaptive MCMC algorithms. *Ann. Appl. Probab. 16*(3), 1462–1505.

Andrieu, C. and J. Thoms (2008). A tutorial on adaptive MCMC. *Statistics and Computing 18*(4), 343–373.

Atchadé, Y. (2010). A cautionary tale on the efficiency of some adaptive Monte Carlo schemes. *Ann. Appl. Probab. 20*(3), 841–868.

Atchadé, Y. and G. Fort (2010). Limit theorems for some adaptive MCMC algorithms with subgeometric kernels. *Bernoulli 16*(1), 116–154.

Atchadé, Y., G. Fort, E. Moulines, and P. Priouret (2011). *Inference and Learning in Dynamic Models*, Chapter Adaptive Markov chain Monte Carlo: Theory and Methods, pp. 33–53. Cambridge University Press.

Baxendale, P. H. (2005). Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Probab. 15*(1A), 700–738.

Bercu, B., P. Del Moral, and A. Doucet (2009). A Functional Central Limit Theorem for a class of Interacting Markov Chain Monte Carlo Methods. *Electron. J. Probab. 14*, 2130–2155.

Brockwell, A., P. Del Moral, and A. Doucet (2010). Sequentially interacting Markov chain Monte Carlo methods. *Ann. Statist. 38*(6), 3387–3411.

Del Moral, P. and A. Doucet (2010). Interacting Markov Chain Monte Carlo methods for solving nonlinear measure-valued equations. *Ann. Appl. Probab. 20*(2), 593–639.

Douc, R. and E. Moulines (2008). Limit theorems for weighted samples with applications to sequential Monte Carlo methods. *Ann. Statist. 36*(5), 2344–2376.

Douc, R., E. Moulines, and J. Rosenthal (2004). Quantitative bounds for geometric convergence rates of Markov chains. *Ann. Appl. Probab. 14*(4), 1643–1665.

Fort, G. and E. Moulines (2003). Polynomial ergodicity of Markov transition kernels,. *Stochastic Processes and Their Applications 103*, 57–99.

Fort, G., E. Moulines, and P. Priouret (2010a). Convergence of adaptive and interacting Markov chain Monte Carlo algorithms. submitted.

Fort, G., E. Moulines, and P. Priouret (2010b). Supplement paper to: Convergence of adaptive and interacting Markov chain Monte Carlo algorithms. submitted.

Fort, G., E. Moulines, P. Priouret, and P. Vandekerkhove (2011). Supplementary paper to "A Central Limit Theorem for Adaptive and Interacting Markov Chains". Technical report, arXiv.

Haario, H., E. Saksman, and J. Tamminen (1999). Adaptive proposal distribution for random walk Metropolis algorithm. *Computational Statistics 14*, 375–395.

Hall, P. and C. C. Heyde (1980). *Martingale Limit Theory and its Application.* New York, London: Academic Press.

Jarner, S. and E. Hansen (2000). Geometric ergodicity of Metropolis algorithms. *Stoch. Process. Appl. 85*, 341–361.

Kou, S. C., Q. Zhou, and W. H. Wong (2006). Equi-energy sampler with applications in statistical inference and statistical mechanics. *Ann. Statist. 34*(4), 1581–1619.

Liang, F., C. Liu, and R. J. Carroll (2007). Stochastic approximation in Monte Carlo computation. *J. Am. Statist. Assoc. 102*, 305–320.

Meyn, S. P. and R. L. Tweedie (2009). *Markov Chains and Stochastic Stability.* London: Cambridge University Press.

Roberts, G. O. and J. S. Rosenthal (2004). General state space Markov chains and MCMC algorithms. *Probab. Surv. 1*, 20–71.

Roberts, G. O. and J. S. Rosenthal (2007). Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms. *J. Appl. Probab. 44*(2), 458–475.

Roberts, G. O. and R. L. Tweedie (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika 83*(1), 95–110.

Rosenthal, J. S. (2009). *MCMC Handbook*, Chapter Optimal Proposal Distributions and Adaptive MCMC. Chapman & Hall/CRC Press.

Saksman, H. and M. Vihola (2010). On the ergodicity of the adaptive Metropolis algorithm on unbounded domains. *Ann. Appl. Probab. 20*, 2178–2203.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* New York: Wiley.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak convergence and empirical processes.* New York: Springer-Verlag.

Wang, F. and D. P. Landau (2001). Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett. 86*, 2050–2053.

LTCI, TELECOM ParisTech & CNRS,
46 rue Barrault,
75634 Paris Cedex 13,
France
E-mail: gersende.fort@telecom-paristech.fr
E-mail: eric.moulines@telecom-paristech.fr

LPMA, Univ. Pierre et Marie Curie,
Boîte courrier 188
75252 PARIS Cedex 05, France,
E-mail: priouret@ccr.jussieu.fr

LAMA, Univ. Paris Est
5 Bd Descartes Champs-sur Marne
77454 Marne-la-Vallée Cedex 2, France.